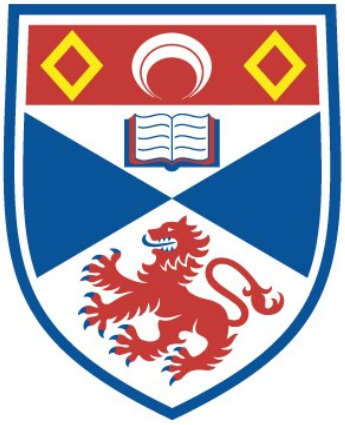
ON THE FAST AND ACCURATE COMPUTER SOLUTION OF PARTIAL DIFFERENTIAL SYSTEMS

Michael T. Hill

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews



1974

DECLARATIONS

I HEREBY DECLARE THAT THIS THESIS HAS BEEN COMPOSED BY MYSELF, THAT THE WORK OF WHICH IT IS A RECORD HAS BEEN DONE BY MYSELF, AND THAT IT HAS NOT BEEN ACCEPTED IN ANY PREVIOUS APPLICATION FOR A HIGHER DEGREE.

MICHAEL T. HILL

THE RESEARCH WAS UNDERTAKEN FULL TIME AT THE VON KARMAN INSTITUTE DURING THE PERIOD OCTOBER 1973 TO DECEMBER 1976, AND THE STUDENT WAS MATRICULATED AS A FULL TIME RESEARCH STUDENT AT THE UNIVERSITY OF ST. ANDREWS FOR THE PERIOD OCTOBER 1974-AUGUST 1979.

PROFESSOR J.J. GINOUX
VON KARMAN INSTITUTE

THE CONDITIONS OF THE RESOLUTION AND REGULATIONS HAVE BEEN FULFILLED

PROFESSOR S.N. CURLE (SUPERVISOR)
DEPT. OF APPLIED MATHEMATICS
UNIVERSITY OF ST. ANDREWS

ON THE FAST AND ACCURATE COMPUTER SOLUTION

OF PARTIAL DIFFERENTIAL SYSTEMS

by

Michael T. Hill

B.Sc. (Hons.) Applied Maths, University of St. Andrews (1973)

Post Graduate Diploma in Fluid Dynamics,
The von Karman Institute for Fluid Dynamics (1974)

Th 9219

TABLE OF CONTENTS

ABSTRACT . . . . . .

1. GENERAL INTRODUCTION 1

1.1 Introduction and historical survey 1

1.2 Introduction to higher order methods 5

2. DESCRIPTION OF THE METHOD 8

2.0 Introduction 8

2.1 Application to ordinary differential equations . . 11

2.2 Some results with the method 25

3. APPLICATION TO PARTIAL DIFFERENTIAL EQUATIONS 27

3.0 Introduction 27

3.1 Parabolic equations 27

3.2 Elliptic equations 37

3.3 Hyperbolic equations 39

3.4 Summary 45

4. FAST METHODS 48

4.0 Discussion 48

5. A RELAXATION METHOD FOR HYPERBOLIC EQUATIONS 51

5.0 Introduction 51

5.1 Relaxation applied to differential equations . . . 53

5.2 Relaxation applied to difference schemes 68

5.2.1 First order differencing 69

5.2.2 Lax's method 74

5.2.3 The Lax-Wendroff method 79

5.2.4 "Euler" differencing 83

5.2.5 A semi-implicit scheme 85

6. CONCLUSIONS 89

REFERENCES 90

ABSTRACT

Two methods are presented for use on an electronic computer for the solution of partial differential systems.

The first is concerned with accurate solutions of differential equations. It is equally applicable to ordinary differential equations and partial differential equations, and can be used for parabolic, hyperbolic or elliptic systems, and also for non-linear and mixed systems. It can be used in conjunction with existing schemes. Conversely, the method can be used as a very fast method of obtaining a rough solution of the system. It has an additional advantage over traditional higher order methods in that it does not require extra boundary conditions.

The second method is concerned with the acceleration of the convergence rate in the solution of hyperbolic systems. The number of iterations has been reduced from tens of thousands with the traditional Lax-Wendroff methods to the order of twenty iterations.

Analyses for both the differential and the difference systems are presented. Again the method is easily added to existing programs.

The two methods may be used together to give one fast and accurate method.

ACKNOWLEDGEMENTS

The author must thank :

the Science Research Council and the von Karman Institute for funding parts of this research, Professor H-J Wirz for supervising the work, and Mme Toubeau for typing so well especially the equations.

1. GENERAL INTRODUCTION

1.1 Introduction and historical survey

The discipline now known as "computational fluid dynamics" appears to have as its date of birth the year 1910, when L.F. Richardson presented his historical paper (1) to the Royal Society. Richardson used as his examples the iterative solution of Laplace's equation, the biharmonic equation and others. His "computers" were boys, being paid at a rate proportional to the number of correct calculations carried out. He combined all of his proposed methods into a large scale practical example, but, because of the small number of time step calculations which were performed, the instability present in one of his procedures was not discovered at the time. This fact serves to highlight the state of affairs in that numerical methods and their corresponding stability analyses must proceed together.

In 1928 Courant, Friedrichs and Lewy published their classical paper (2) in which they established certain existence theorems and uniqueness theorems for partial differential systems. Although the authors were primarily interested in using finite difference formulations as a tool for pure mathematics, their work has since become the "cornerstone" for modern practical finite difference solutions.

Until the arrival of the electronic computer in the 1940s, the emphasis in numerical methods had been on the "jury-

type" elliptical problems (where the problem is solved over the whole field simultaneously). The first study of the numerical solution of a viscous fluid dynamics problem was presented by Thom (3) in 1933. He studied the viscous flow around-circular cylinder, for which some analytical solutions are possible for comparison.

In 1938 Shortley and Weller (4) presented what was basically an improved version of Richardson's method, using over relaxation. They also included, for the first time, an identification and analysis of the convergence rates.

During the Second World War, J. von Neumann and others at the Los Alamos Scientific Laboratory in the USA had done much work on the development of the first electronic computers, and their application initially to ballistic problems. A considerable amount of effort was given to the consideration of convergence, numerical stability and the uniqueness of the solutions. Much of the work carried out during the war was classified as secret, and, in 1946, Southwell (5) presented a relaxation method which clearly fulfilled two aims. Firstly, it obtained a better rate of convergence, and secondly, it succeeded in making the work more interesting for the human computers. This latter was because they had to scan the computational mesh for the largest residual(s) and update the solution accordingly. In fact, this advantage becomes a disadvantage when the method is applied on electronic computers, because scanning the grid of mesh-points would take longer then the arithmetic involved. Thus, electronic digital computers provided an incentive for the further development

of the Liebmann variation of Richardson's method. In 1950 Frankel (6) presented the method now known as "successive over relaxation" which is still widely used for elliptic problems.

As the electronic computer begun to become available so the centre of interest moved from elliptic (usually steady-state) problems to parabolic (time dependent) problems. This was because it became feasible to attempt time development problems. The most well known of the many parabolic methods to be published in the 1940s was in the Crank-Nicolson paper (7) published in 1947. Although this (Crank-Nicolson method) is still used in many methods published at the present time, it uses a very simple formula for calculating the next time-like step, and is perhaps better suited to human computers than to electronic digital computers. This situation may become reversed when parallel processing becomes widely available.

In the early 1950s, the wartime work carried out at Los Alamos began to be published, including von Neumann's (8) famous criterion for stability of parabolic finite difference equations, and a method of analysing a linearised system.

In 1955 and 1956, Peaceman & Rachford (9) and Douglas & Rachford (10) presented their alternating-direction-implicit (ADI) method for parabolic systems of equations which allowed an arbitrarily large time step for stability (but, of course, not for accuracy).

Although the fundamental paper dealing with hyperbolic systems had been presented as early as 1928 (by Courant, Friedrichs and Lewy), little work appears to have been done on them until the mid 1950s. The 1928 paper presented a necessary stability criterion, that the finite difference domain of dependence must include the continuous (i.e., differential) domain of dependence The simplest and earliest method for hyperbolic equations is that of Lax (11) in 1954. This was improved in the Lax-Wendroff (12) method (1960), improved again in two step methods such as Richtmyer's (13) (1963) and MacCormack's (14) (1969). Lax's 1954 paper also presented an argument for writing the system in conservation or divergence form - that is returning to the physics of the problem and writing the equations by analogy to Newton's law of conservation of mass, momentum and energy rather than expanding the terms to a "neat" mathematical formulation. This form is especially important in calculations involving shocks.

Taylor (15) remarks that the majority of the current work done on hyperbolic systems is being carried out at Los Alamos, where "almost unlimited resources are available". These include the Particle In Cell (PIC) method of Harlow and others (16) and the Explosive In Cell (EIC) method of Mader (17) (1964). These and other methods include treatment of fluid / fluid boundaries as well as discontinuities.

Only these latest methods have dealt with equations which make use of the power of the computer, rather than just purely putting well proven methods onto the computer to do the mundane and methodical calculations more rapidly. The two methods

presented herein use the computer to a greater benefit during the solution of the system.

The method presented in the first part of this paper allows the computer to solve a given partial differential system (where system is understood to include a set of partial differential equations and the necessary and sufficient boundary conditions) to any required degree of accuracy. Part of this description has been published previously by the current author (18). In the second part of this paper a method is proposed which will solve hyperbolic partial differential systems up to two orders of magnitude faster than existing methods. The obvious combination of the two methods therefore presents a very fast and very accurate solution of partial differential systems.

1.2 Introduction to higher order methods

Higher order methods (HOMs) are of interest not only as such, but also because of the fact that an improvement in accuracy will lead to a situation where larger grid steps can produce the same (lower) accuracy. Therefore a balance between accuracy and grid size will lead to an optimised method.

For elliptic equations, classically second order differencing is considered adequate, giving(in two dimensions) a simple five point difference formula. (The "Mehrstellen" methods have improved this to a certain extent). On a unit square, in order to solve Laplace's equation to a normalised accuracy

of 10^{-6}, a grid of 1000 × 1000 points is required, that is to say, a mesh size of 10^{-3}. If each iteration on each grid point requires only 3 multiplications (ignoring additions), that gives a total of of 3×10^6 per iteration, and so 33 iterations give an unreasonable 10^8 multiplications. A fourth order method would require a grid of 30 × 30 points, a grid size of .03 giving 900 grid points and 89.000 multiplications for 33 iterations. Clearly a sixth order method needs only a 10 × 10 grid, 300 multiplications per iteration or 10.000 for 33 iterations. Clearly this would reduce the computation time from hours to seconds.

The solution is unfortunately not as simple as this. One of the problems encountered in using higher order methods is that the majority of them need additional boundary conditions. This fact often proves to be a major hurdle, as these additional boundary conditions must be compatible with the (as yet unknown) solutions. This is a numerical effect and usually has little to do with the physics of the problem - save in the conservation of matter, momentum and energy. This effect can easily be demonstrated by considering a simple example. The equation

$$u_t = u$$

has a solution :

$$u = u_0 \, e^t.$$

Only one boundary condition is needed in order to quantify u_0, (which is normally considered as an initial value). When solved "numerically" with a first order method, again one condition is required. The solution of the simple difference equation

$$u_{j+1} - u_j = \Delta t\ u_j$$

is

$$u_{j+1} = u_j\ (1 + \Delta t)$$

and once again if u_0 is given then the problem is completely solved. When solved numerically with a second order method such as a centred difference, then the difference equation is

$$u_{j+1} - u_{j-1} = 2\Delta t\ u_j$$

which would require both u_0 and u_1 in order to provide a complete solution (or any two equivalent conditions). In general, with an Nth order differential equation and an Mth order difference method (M-N) extra conditions must be conjured up. The method presented herein does not have this drawback.

2. DESCRIPTION OF THE METHOD

2.0 Introduction

In this section a method is described for achieving a numerical solution of a differential equation to any order of accuracy. It is capable of giving a degree of accuracy bounded only by the accuracy of the computer used, if necessary, but is more likely to be used as a method of obtaining fourth or sixth order accuracy in place of first or second order previously. A corollary of a higher order method is a method yielding the same (low) accuracy with fewer nodal points - that is with less computational effort.

The method is easily applied to all types of partial differential equations, but in order to make its utilisation clear, first its use is demonstrated by solving an ordinary differential equation.

To avoid the use of a computer initially, a simple linear second order ordinary differential equation is taken, which has a well known algebraic solution. This equation can be reduced in discrete variables to a simple linear second order difference equation, also with known solution.

Consider the second order linear harmonic ordinary differential equation

$$\frac{d^2f}{dx^2} + \pi^2 f = 0 \qquad (1)$$

together with the boundary conditions

$$f(x=0) = 0$$
$$f(x=\tfrac{1}{2}) = 1 \qquad (2)$$

This equation can easily be shown to possess the exact solution

$$f(x) = A \sin\pi x + B \cos\pi x \qquad (3a)$$

which the boundary conditions then reduce to

$$f(x) = \sin\pi x \qquad (3b)$$

In order to solve the system (1) and (2) numerically, a difference representation must be found for the second derivative, $\frac{d^2f}{dx^2}$. The most widely used discrete form is obtained by expanding f in a Taylor series about some point x at intervals h, as follows

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2} f''(x) + \frac{h^3}{6} f'''(x) + O(h^4) \qquad (4)$$

$$f(x-h) = f(x) - hf'(x) + \frac{h^2}{2} f''(x) - \frac{h^3}{6} f'''(x) = O(h^4) \qquad (5)$$

where primes denote differentiation with respect to x.
Adding these two expressions and subtracting 2f(x) yields

$$f(x+h) - 2f(x) + f(x-h) = h^2 f''(x) + O(h^4) \qquad (6)$$

We now introduce some notation for the discrete variables. Let

$$x_j = jh \qquad j = 0,1,\ldots,N \tag{7}$$

$$f_j = f(x_j) \tag{8}$$

$$\delta_x^2 f_j = f(x_j+h) - 2f(x_j) + f(x_j-h) \tag{9}$$

Using this notation, our original differential equation (1) can now be written as a second order difference equation, namely

$$\delta_x^2 f_j + h^2\pi^2 f_j = 0 \tag{10}$$

correct to order h^2.

The difference equation (10) is also linear, and by comparison with (3a) can be shown to have the solution

$$f_j = A\sin\pi\lambda_j + B\cos\pi\lambda_j \tag{11}$$

and on applying the boundary conditions

$$f_0 = 0 \qquad f_N = 1 \tag{12}$$

the solution becomes

$$f_j = A \sin\pi\lambda_j \tag{13}$$

$$A = \frac{1}{\sin(\frac{\pi\lambda}{2h})}$$

$$\lambda = \frac{2}{\pi} \sin^{-1} (\frac{\pi h}{2}).$$

Expanding λ in a power series in h gives :

$$\lambda = \frac{2}{\pi}\left[\frac{\pi h}{2} + \frac{\pi^3 h^3}{24} + \frac{\pi^5 h^5}{480} + O(h^7)\right] \tag{14}$$

$$= h\left[1 + \frac{\pi^2 h^2}{12} + \frac{\pi^4 h^4}{240} + O(h^6)\right] \tag{15}$$

It can immediatley be seen that the solution (13) is

$$f_j = \frac{1}{\sin\frac{\pi}{2}\left(1+O(h^2)\right)} \sin\pi hj \left[1+O(h^2)\right] \tag{13a}$$

That is to say, the solution is formally accurate to second order in h. The most usual way of expressing this fact is to write

$$f(x) = f_j + O(h^2)$$

but instead of an additive order function a multiplicative function can be defined by writing $f(x) = \alpha f_j$, where $\alpha = 1 + O(h^2)$. This idea forms a basis for the higher order method.

2.1 Application to ordinary differential equations

In order to improve the accuracy of the solution, that is to make the numerical solution closer to the exact solution (3), there are several options available. More points can be taken in the difference expression (9), and this is perhaps the most common tool used. Multistep methods such as the Runge-Kutta scheme, and combinations of the continuous and discrete formulations are also used. Here, rather than doing this, we return to the formula for the second derivative of f.

Consider a new approximation to $\frac{d^2f}{dx^2}$ given by

$$\frac{d^2f}{dx^2} = \alpha \frac{\delta_x^2 f}{h^2} \tag{16}$$

where the coefficient α has yet to be determined and will probably be a function of x.

Rewriting the difference equation (10) by using the definition (16) then leads to the following equation :

$$\delta_x^2 f_j + \frac{h^2\pi^2}{\alpha} f_j = 0 \tag{17}$$

This difference equation also possesses an exact solution, namely

$$f_j = a \sin\pi\mu j + b \cos\pi\mu j \tag{18}$$

and then upon applying the boundary conditions (12), this reduces to :

$$f_j = a \sin\pi\mu j$$

$$a = \frac{1}{\sin \frac{\pi\mu}{2h}} \tag{19}$$

$$\mu = \frac{2}{\pi} \sin^{-1} \left(\frac{\pi h}{2\sqrt{\alpha}}\right)$$

This solution is similar to (13); in fact, it is the same if $\alpha = 1$. With a different value for α we can, naturally, obtain different solutions, and with the "right" value this can be the exact solution of the ordinary differential equation (1).

The outstanding problem is how to evaluate α to give this exact solution. By choosing α we are able to ensure that the solution (19) is the same *at the point* x_j as the solution (3b) of equation (1). By comparison it can be seen that the choice should be such that

$$a = 1 \qquad \text{that is} \qquad \sin\left(\frac{\pi\mu}{2h}\right) = 1 \tag{20}$$

which reduces to

$$\mu = h \tag{21}$$

and hence

$$\mu = \frac{2}{\pi} \sin^{-1}\left(\frac{\pi h}{2\sqrt{\alpha}}\right) = h \tag{22}$$

leading to

$$\alpha = \left(\frac{\frac{\pi h}{2}}{\sin \frac{\pi h}{2}} \right)^2 \tag{23}$$

It should be noted here that in the limit as $h \to 0$, then $\alpha \to 1$, and hence this formulation is consistent.

Truncating the power series in h for α after a given number of terms yields a different accuracy. For example, for a given value of h, by using respectively

$$\alpha = 1 \tag{24}$$

$$\alpha = 1 + \frac{\pi^2 h^2}{12} \tag{25}$$

$$\alpha = 1 + \frac{\pi^2 h^2}{12} + \frac{\pi^2 h^2}{240} \qquad (26)$$

a solution correct to order h^2, h^4 and h^6 can be calculated.

As an illustration of this fact, figure 1 is presented. This figure shows a plot of log($k\pi h$) (horizontally) versus log(max|error|), where k is the frequency. The four curves represent the values obtained for second, fourth, sixth and eighth order methods. The figure clearly demonstrates that the error is, in fact, proportional to h to the power 2, 4, 6 or 8 respectively, as is shown by the gradient of the lines. By means of this figure the necessary grid mesh-size to give some prescribed error at a frequency k can be determined.

By contrast, figure 2 shows the same plot of log($k\pi h$) versus log(max|error|), but on this occasion the solution was evaluated using a CII Mitra 15 computer, whereas the data for figure 1 was calculated using a Control Data 6500. (The Mitra has four or five digit accuracy. By comparison, the CDC carries about 15 significant digits). This illustrates clearly that, of course, the method is limited by the accuracy of the computer used. Second, fourth, sixth or eighth order accuracy can be achieved provided that this is within the machine accuracy. In this case, no solution can be obtained with an error better than 10^{-7}.

In general, the differential equation will not possess such a simple solution, and then neither will the difference

Key to figures 1 and 2

Order	k = 1	k = 5	k = 9
2	A	B	C
4	D	E	F
6	G	H	J
8	K	L	M

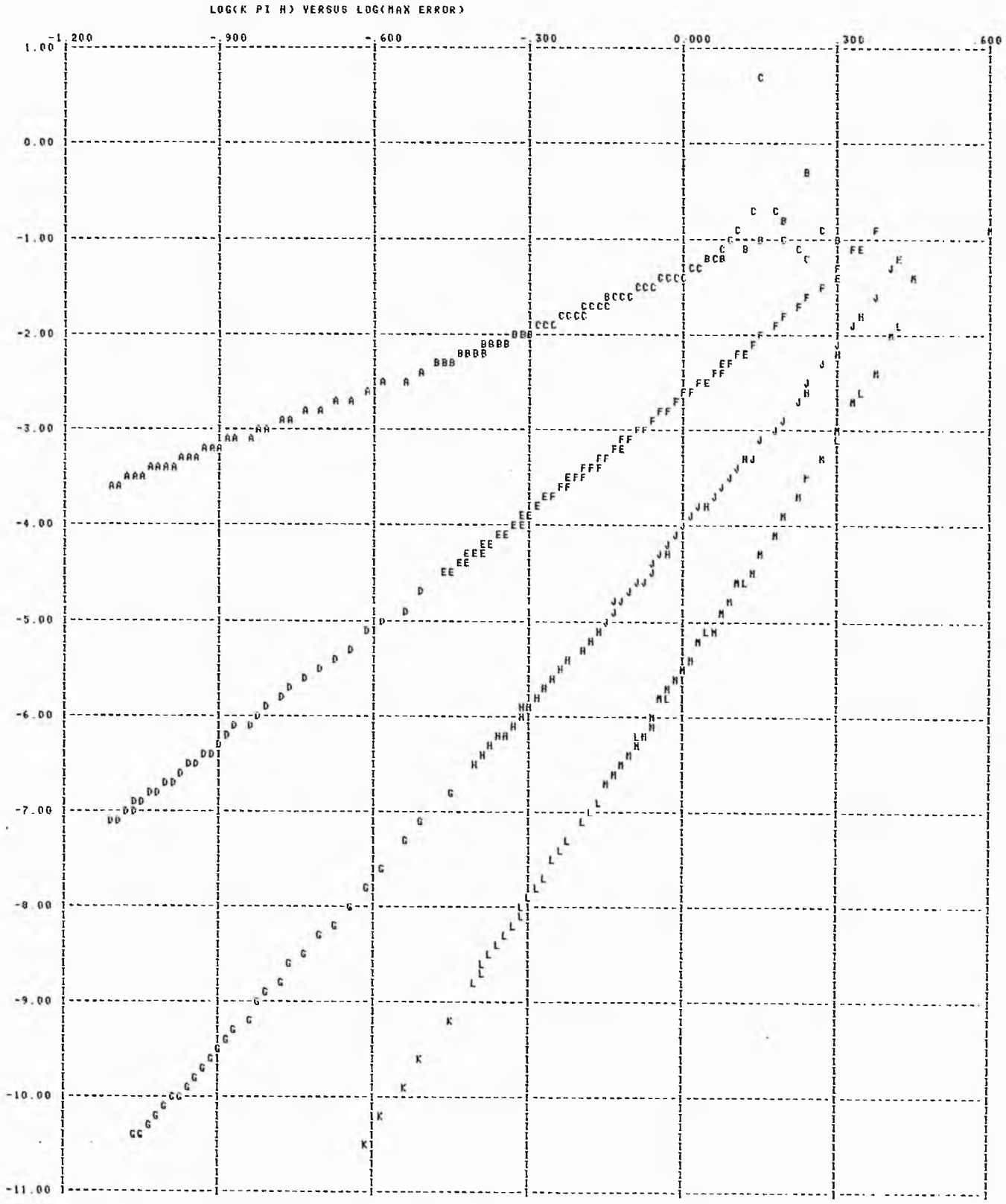

LOG(K PI H) VERSUS LOG(MAX ERROR)
-1.200
-.900
-.600
-.300
0.000
.300
.600
1.00
0.00
-1.00
-2.00
-3.00
-4.00
-5.00
-6.00
-7.00
-8.00
-9.00
-10.00
-11.00

FIGURE 1

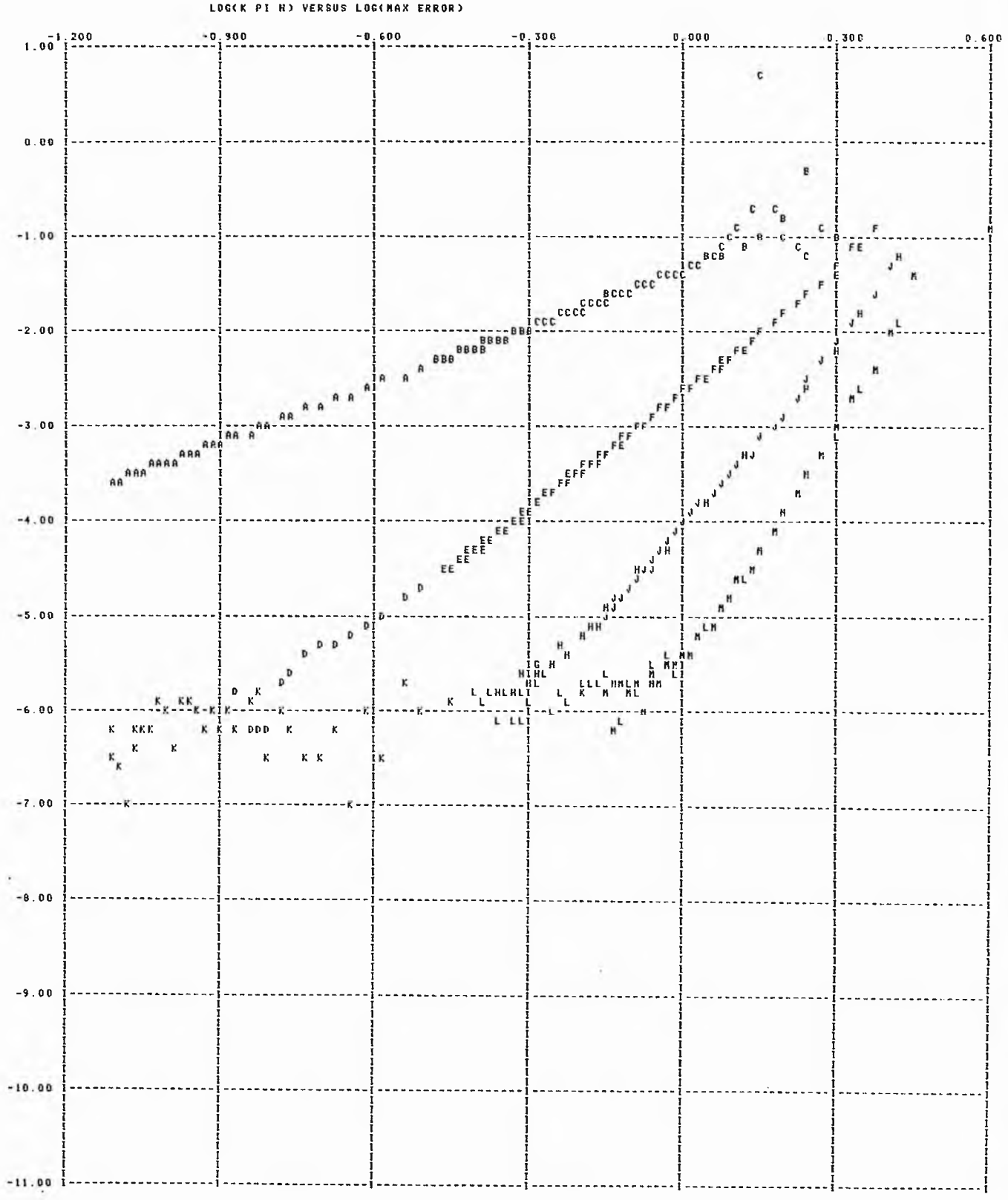

LOG(K PI H) VERSUS LOG(MAX ERROR)
-1.200
-0.900
-0.600
-0.300
0.000
0.300
0.600
1.00
0.00
-1.00
-2.00
-3.00
-4.00
-5.00
-6.00
-7.00
-8.00
-9.00
-10.00
-11.00

FIGURE 2

equation possess a simple solution. We would then resort to solving the equation numerically by means of an electronic digital computer. That this method is ideally suited to solution by computer will become obvious in what follows.

It will be recognised that, so far, this procedure only seems to be useful when the solution is already known. This is not the case, and some alternative methods for calculating α are developed below. If the solution is defined on some given interval, or is known to be a harmonic function, Fourier analysis can be helpful. The discrete Fourier transform is defined as follows. Given a function $f(x)$ which is periodic on an interval $0 < x < X$, where the interval is divided into $N+1$ points such that $Nh = X$, then

$$f(x_j) = a_0 + \sum_{n=1}^{N/2} \left[a_n \cos 2\pi n x_j + b_n \sin 2\pi n x_j \right] \qquad (27)$$

where $x_j = (j-1)h \qquad j = 1,\ldots,N+1$

$$a_0 = \frac{1}{N} \sum_{j=1}^{N} f(x_j) \qquad (28a)$$

$$a_n = \frac{2}{N} \sum_{j=1}^{N} f(x_j) \cos 2\pi n x_j \qquad n = 1,\ldots,\frac{N}{2} - 1 \qquad (28b)$$

$$a_{N/2} = \frac{1}{N} \sum_{j=1}^{N} f(x_j) \cos 2\pi \frac{N}{2} x_j \qquad (28c)$$

$$b_n = \frac{2}{N} \sum_{j=1}^{N} f(x_j) \sin 2\pi n x_j \qquad n = 1,\ldots,\frac{N}{2} \qquad (28d)$$

Now, by rewriting equation (16) as

$$\alpha = \frac{h^2 \frac{d^2 f}{dx^2}}{\delta_x^2 f_j} \tag{29}$$

with a Fourier series expansion for f, it is possible to evaluate the numerator of this expression. Hence, it is possible to evaluate α. The second derivative can be approximated as the limit as $h \to 0$, namely

$$\left.\frac{d^2 f}{dx^2}\right|_j \simeq \lim_{h\to 0} \frac{\delta_x^2 f_j}{h^2} = -(2\pi)^2 \sum_{n=1}^{N/2} n^2 (a_n \cos 2\pi n x_j + b_n \sin 2\pi n x_j) \tag{30}$$

and

$$\delta_x^2 f_j = \sum_{n=1}^{N/2} (\cos 2\pi n h - 1)(a_n \cos 2\pi n x_j + b_n \sin 2\pi n x_j) \tag{31}$$

Hence,

$$\alpha = (2\pi h)^2 \frac{\sum_{n=1}^{N/2} n^2 (a_n \cos 2\pi n x_j + b_n \sin 2\pi n x_j)}{\sum_{n=1}^{N/2} (1-\cos 2\pi n h)(a_n \cos 2\pi n x_j + b_n \sin 2\pi n x_j)} \tag{32}$$

If $a_n = 0$ for all n except a_p, and $b_n = 0$ for all n,

$$\alpha = \frac{(2\pi h)^2 p^2 a_p \cos 2\pi p x_j}{(1-\cos 2\pi p h) a_p \cos 2\pi p x_j}$$

$$= \frac{(2\pi h p)^2}{1-\cos 2\pi p h} = \left[\frac{(\pi h p)}{\sin \pi h p}\right]^2 \tag{33}$$

which is of a similar form to equation (33).

If we have more than one of the coefficients a_n, b_n non zero, then α will be dependent on these coefficients and it is therefore necessary to find a way to compute these. One such a way is to solve the difference equation setting $\alpha = 1$ to get a first approximation to f_j of order h^2. Then use this f_j to calculate a better approximation to α in equation (32) repeating this process if necessary.

In the case where the solution sought is known not to be periodic, or nothing is known about the harmonic properties of the solution, then the above Fourier analysis may not be helpful, and then some other means for evaluating α should be used.

Returning to equation (29) for α suggests that α can be determined if the second derivative in the numerator can be evaluated. Using a process similar to that described above, if we first solve the difference equation obtained by setting $\alpha = 1$, we will have some knowledge of the form of the solution, it is quite feasible to use a higher order expression for $\left.\frac{d^2f}{dx^2}\right|_j$, for example

$$\left.\frac{d^2f}{dx^2}\right|_j = \frac{1}{12h^2}\left\{-(f_{j+2}+f_{j-2})+16(f_{j+1}+f_{j-1})-30f_j\right\} + O(h^4) \quad (34)$$

Then substituting this expression into equation (29) produces a second approximation to α which can then be used to produce a second approximation to f_j. Again, this procedure can be repeated until no difference is obtained between successive

values of f_j. It appears, in the various cases tried to date, that three or four iterations prove adequate. A non centred 5 point difference would be used at j = 1 and j = N - 1.

Using matrix-operator notation, and superscripts in parentheses to denote the iteration level, this procedure can be written briefly as follows. Let

$$(T\underline{f})_j = \delta_x^2 f_j \tag{35}$$

$$(P\underline{f})_j = \left.\frac{d^2 f}{dx^2}\right|_j \qquad \text{e.g. as given by equation (34)} \tag{36}$$

Equation (5) can be written as

$$T\underline{f} + h^2\pi^2 I\underline{f} = \underline{b} \tag{37}$$

where I is the unit matrix, and $\underline{b}$ contains the boundary conditions. If

$$T_1 = T + h^2\pi^2 I \tag{38}$$

then equation (37) becomes

$$T_1\underline{f} = \underline{b} \tag{39}$$

The iterative scheme for $\underline{f}$ will then be :

$$T_1\underline{f}^{(1)} = \underline{b} \tag{40}$$

$$D^{(1)}\, T_1 f^{(1)} = P\underline{f}^{(1)} \tag{41}$$

$$D^{(1)}\, T_1\underline{f}^{(2)} = \underline{b} \tag{42}$$

$$\vdots$$

$$D^{(k)} \; T_1 \underline{f}^{(k)} = P\underline{f}^{(k)} \tag{43}$$

$$D^{(k)} \; T_1 \underline{f}^{(k+1)} = \underline{b} \tag{44}$$

where equations (40), (42) and (44) are used to solve for $\underline{f}$ and then equations (41) and (43) are used to evaluate D, which is the diagonal matrix with elements α_j.

If, in this sequence $\{f^{(k)}\}$ and correspondingly $\{D^{(k)}\}$ k is a number larger than 4, then some acceleration technique should be applied. For instance, an over relaxation expression of the form

$$D^{(k)} = (1-\omega) \; D^{(k)} + \omega d^{(k)}$$

where $d^{(k)}$ is the result of solving an equation of the form (43) for D.

It is straightforward to show that, if the sequence of iterates $\{f^{(k)}\}$ tends to a limit $\underline{f}^*$ say, then this limit is a higher order solution. Assuming that a limit exists, then equations (43) and (44) become

$$D^* \; T_1 \underline{f}^* = P\underline{f}^* \tag{45}$$

$$D^* \; T_1 \underline{f}^* = \underline{b} \tag{46}$$

and hence

$$P\underline{f}^* = \underline{b} \tag{47}$$

that is, the solution of the iterative procedure is the solution

of equation (1) using a fourth order approximation for the second derivative. For example, if P is chosen to use all the $N = \frac{1}{2h}$ points of the solution, then the solution $\underline{f}$ will be Nth order correct in h. If the solution is a polynomial of order m, where $m < N$, then the solution will be exact.

The extension to include first order derivatives is similar. To solve the equation

$$\frac{d^2f}{dx^2} + 2\ell \frac{df}{dx} + kf = 0 \tag{48}$$

a second order difference for the first derivative is introduced with an unknown coefficient, β,

$$\left.\frac{df}{dx}\right|_j = \beta \frac{f_{j+1} - f_{j-1}}{2h} \tag{49}$$

or if D_2 is the diagonal matrix with elements β, and T_2 is the operator :

$$(T_2\underline{f})_j = f_{j+1} - f_{j-1} \tag{50}$$

then equation (48) becomes in discrete form

$$T_1 f + h\ell\, T_2\underline{f} + h^2\, kI\underline{f} = \underline{b} \tag{51}$$

or if

$$T_1 + h\ell\, T_2 + h^2\, kI = T_3 \tag{52}$$

$$T_3\underline{f} = \underline{b} \tag{53}$$

The equation corresponding to equation (41) is

$$D_2^{(1)} \; T_2 \underline{f} = P_2 \underline{f} \tag{54}$$

leading to the iterative system

$$T_3 \underline{f}^{(1)} = \underline{b} \tag{55}$$

$$D^{(1)} \; T_1 \underline{f}^{(1)} = P \underline{f}^{(1)} \tag{56}$$

$$D^{(1)} \; T_2 \underline{f}^{(1)} = P_2 \underline{f}^{(1)} \tag{57}$$

$$\left\{ D^{(1)} \; T_1 + h\ell \; D_2^{(1)} \; T + h^2 \; kI \right\} \; f^{(2)} = \underline{b} \tag{58}$$

and so on. Here P_2 is some higher order difference operator for the first order derivative.

The advantages of using such a scheme are immediately clear. To solve equation (55) for $\underline{f}$ requires the inversion of the matrix T , which is a tri-diagonal matrix, i.e., the only non zero elements are the diagonal and one element to each side of the diagonal. There exists a well known algorithm for inverting such a matrix, which requires much less work than inverting the matrix P which has more non zero elements, and may even have no zero entry.

A further and important advantage of this scheme is that, although the accuracy of the solution may be fourth order or higher, the method only requires two boundary conditions, as would be expected for a second order differential equation. It

is more often the case that higher order methods require more boundary conditions than the differential equation, and determining these extra boundary conditions satisfactorily can involve a disproportionately large amount of work.

2.2 Some results with the method

To show how quickly the iterative procedure outlined above converges to the high order solution, the following examples were used :

1) with $\alpha = 1$, $h = 0.05$, a five point formula such as equation (45), gives immediately ; $\alpha = 1.00205$,
   which is the correct value as given by equation (23);
2) with $\alpha = 1$, $h = 0.01$, equation (45) gives a value :
   $\alpha = 1.000082$
   which again is correct and so there is no need for iteration;
3) taking α too small, namely, $\alpha = 0.1$ and $h = .05$, only two iterations are required. The second order solution gives :
   $\alpha = 1.020562$
   first iteration yields
   $\alpha = 1.002015$
   and a second iteration results in
   $\alpha = 1.002052$
   as example (1);
4) taking α too large, $\alpha = 10$, with $h = .05$ as in example (3), gives on successive iterations :
   $\alpha = 1.000206$, $\alpha = 1.002056$, $\alpha = 1.002052$.

These four examples show clearly, provided the value of α is such that μ exists in equation (22), namely $\left|\frac{\pi h}{2\sqrt{\alpha}}\right| \leq 1$, that any value for α causes the procedure to converge rapidly to a high order solution.

With relaxation these sequences can be made to converge even more rapidly.

3. APPLICATION TO PARTIAL DIFFERENTIAL EQUATIONS

3.0 Introduction

The extension of this approach to the higher order solution of partial differential equations is reasonably simple. Instead of our multiplicative variables α etc being simple constants or functions of one variable, they will now be more complicated functions of two or more variables, and we will have one such coefficient for each derivative.

For the sake of clarity, the three basic types of partial differential equations will each be treated separately to show some of the features peculiar to each and how our higher order method copes with them. The first type treated is parabolic equations.

3.1 Parabolic equations

As an example of a parabolic equation, consider the linearized form of the Burgers equation :

$$u_t + Uu_x = u_{xx} \tag{59}$$

where subscripts t and x denote partial differentiation, U is a given function of x and t, and some boundary and initial conditions are prescribed. Writing

$$u_{(x,t)} = u(jh,k\tau) = u_j^k \tag{60}$$

$$u_t = \gamma \frac{u_j^{k+1} - u_j^k}{\tau} + O(\tau) \tag{61}$$

$$u_x = \beta \frac{u_{j+1}^k - u_{j-1}^k}{2h} + O(h^2) \tag{62}$$

$$u_{xx} = \alpha \frac{u_{j+1}^k - 2u_j^k + u_{j-1}^k}{h^2} + O(h^2) \tag{63}$$

where α, β and γ are unknown functions of x and t, leads to the difference equation for u_j^{k+1} :

$$u_j^{k+1} = u_j^k - U(x_j,t_k) \frac{\beta}{\gamma} \frac{\tau}{2h} (u_{j+1}^k - u_{j-1}^k)$$

$$+ \frac{\alpha}{\gamma} \frac{\tau}{h^2} (u_{j+1}^k - 2u_j^k + u_{j-1}^k) \tag{64}$$

Setting the coefficients α, β and γ to unity will lead to a solution for u correct to order $\tau + h^2$. This solution can then be used to determine new values of these coefficients α, β and γ to obtain a better approximation to the solution of equation (59), following the procedure used above in section 2.2.

In order to understand how this procedure is carried out, consider equation (59) with $U(x,t) \equiv 0$, which is then the well known and well studied diffusion equation. The form of differencing described above by equations (61), (62) and (63) then leads to the explicit difference equation

$$u_j^{k+1} = u_j^k + \frac{\alpha}{\gamma} \frac{\tau}{h^2} (u_{j+1}^k - 2u_j^k + u_{j-1}^k) \tag{65}$$

If we now prescribe the following initial and boundary conditions

$$\begin{aligned} u(x,0) &= x + \sin 2\pi\omega x \\ u(0,t) &= 0 \\ u(1,t) &= 1 \end{aligned} \tag{66}$$

then the exact solution of the differential equation

$$u_t = u_{xx} \tag{67}$$

is

$$u(x,t) = x + e^{-4\pi^2\omega^2 t} \sin 2\pi\omega x \ . \tag{68}$$

The exact solution of the difference equation (65) subject to the initial and boundary conditions (66) is

$$u_j^k = jh + \zeta^k \sin 2\pi\omega jh \tag{69}$$

where

$$\zeta = 1 - 4s \sin^2 \pi\omega h \tag{70}$$

and

$$s = \frac{\tau}{h^2} \frac{\alpha}{\gamma} \ . \tag{71}$$

Comparing equations (68) and (69) it can be seen that the exact value for ζ is

$$\zeta_{ex} = e^{-4\pi^2\omega^2\tau} \ . \tag{72}$$

To give these variables some numbers, let

$$h = 0.1; \qquad \tau = 0.004; \qquad \omega = 1.$$

Then, with $\alpha = \gamma = 1$,

$$\zeta = 1 - 0.4 \sin^2 0.1\pi = 0.84721359$$

which should be compared with

$$\zeta_{ex} = 0.85392350.$$

Using a three point formula for the numerator of γ (correct to order $\zeta^2 = 0.000016$) and a five point formula for the numerator of α (as in the definition of α, equation (29)), $\gamma = 1.0901700$ and $\alpha = 1.0318305$, giving

$$\zeta_{\alpha,\gamma} = 0.85538982$$

which is a better value.
Repeating this procedure gives

$$\zeta_{\alpha,\gamma} = 0.85463764$$

which is a further improvement, and already correct to order $h^4 + \tau^2$.

At the end of the section above on ordinary differential equations it was pointed out that, whereas most other higher order methods need extra boundary conditions, this method requires no auxiliary conditions. The scheme uses only those conditions which are given as input to the problem. In general, with an n^{th} order difference for an m^{th} order derivative, an extra (n-m) condition must be artificially defined. In many cases the satisfactory definition of these non-physical boundary conditions can require as much work as the solution of the rest of the problem. An arbitrary choice can often render a well posed problem ill posed.

As an example of this, consider again equation (59)

$$u_t + Uu_x = u_{xx} \qquad (59)$$

We have a first derivative with respect to t, a first and a second with respect to x. Thus we need one condition (in this case it is an initial condition) in t and two in x. The difference formulation (65) has a first difference with respect to k, and a second difference with respect to j. Therefore, it requires one initial condition and two boundary conditions, the same as the original differential formulation. If, on the other hand, for a more accurate solution, second order differencing for t and fourth order for x are used, then two initial conditions and four spatial boundary conditions are required. However, the statement of the problem provides only one initial and two boundary conditions, so the remaining conditions must be invented. With the present higher order scheme, the equation is written as in equation (65) exactly as if the formulation were first order in time and second order in space. The effect of this is that precisely the same number of conditions is required for the difference system as for the differential system. The same is true for equations of elliptic and hyperbolic type. This is obviously a very valuable side effect of the method. When extra (auxiliary) boundary conditions are added indiscriminately then the problem can soon become ill posed, with the result that the solution 'blows up' or diverges, even when the continuous solution is bounded.

Another useful effect of this method is the following. If the difference in the denominator of an expression like equa-

tion (29) is evaluated at time level k + 1, then the resulting system is 'implicit'. All calculations are done using an explicit scheme, but the converged solution will be as if the scheme were implicit. Thus the scheme becomes independent of time-like step-size. An additional advantage of an implicit scheme is that it adds in the space-like coordinate conditions at the new level immediately rather than working out the interior points independently of the boundary conditions and including them afterwards.

Some results of all these effects are shown in some examples. Tables 1, 2 and 3 illustrate the rapid increase obtained by the application of this method to the diffusion in different forms. The five rows of these tables are :

1. the computed values of the solution obtained by applying the standard second order in space, first order in time, implicit difference and proceeding to a time t = 0.4;
2. the exact solution at time t = 0.4;
3. the absolute value of the error, that is the difference between lines 1 and 2;
4. the values given by calculating the coefficients and correcting the solution at time t = 0.4;
5. the remaining error, the difference between line 2 and line 4.

Table 1 shows the results for the diffusion equation $u_t = u_{xx}$. The boundary conditions applied are such that the solution is defined at each boundary (Dirichlet problem). It can be seen that line 1 is second order correct in h, and line 5 is fourth order correct in h.

TABLE 1

$u_t = u_{xx}$ $\quad u_{(t=0)} = x+\sin\pi x$ $\quad u_{(x=0)} = 0$ $\quad u_{(x=1)} = 1$

Exact solution $u(x,t) = x+e^{-\pi^2 t}\sin\pi x$ $\quad \Delta t = \Delta x = .1$ implicit method

	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.
$u^{1}_{.4}$	.120152	.238331	.352758	.462021	.565213	.662021	.752758	.838331	.920152	1.
exact $u^{}_{.4}$	.105963	.211342	.315611	.418351	.519296	.618351	.715611	.811342	.905963	1.
$\lvert\varepsilon\rvert$	.014189	.026989	.037146	.043669	.045916	.043669	.037146	.026989	.014189	0.
corrected $u^{}_{.4}$	.105663	.210774	.314830	.417434	.518331	.617434	.714830	.810774	.905663	1.
$\lvert\varepsilon\rvert$	.000300	.000568	.000781	.000917	.000917	.000917	.000781	.000568	.000300	0.

T A B L E 2

$u_t = u_{xx} - 2$ $\quad u_{(t=0)} = x^2 + \sin\pi x$ $\quad u_{(x=0)} = 0$ $\quad u_{(x=1)} = 1$

Exact solution $u(x,t) = x^2 + e^{-\pi^2 t} \sin\pi x$ $\quad \Delta t = \Delta x = .1$ implicit method

	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.
$u^1_{.4}$	.0301518	.0783310	.142758	.222021	.315213	.422021	.542758	.678331	.830152	1.0
exact $u_{.4}$	.0159629	.0513421	.105611	.178352	.269296	.378352	.505611	.651342	.815963	1.0
$\lvert\varepsilon\rvert$	.0141889	.0269889	.037147	.043669	.045917	.043669	.037147	.0269889	.0141889	0.0
corrected $u_{.4}$	.0158510	.0511299	.105319	.178010	.268937	.378010	.505319	.651130	.815851	1.0
$\lvert\varepsilon\rvert$	.000112	.000212	.000292	.000342	.000359	.000342	.000292	.000292	.000112	0.0

TABLE 3

$u_t = u_{xx} - 2 \qquad u_{(t=0)} = 0 \qquad u_{(x=0)} = 0 \qquad \frac{\partial u}{\partial x}(x=1) = 0$

Exact solution $u(x,t) = x^2 - 2x - \frac{32}{\pi^3} \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)^3} e^{-\frac{(2n+1)^2\pi^2 t}{4}} \cos \frac{1}{2}(2n+1)(x-1)$

Steady state $u = x^2 - 2x$ reached at $t = 8.0$ $\qquad \Delta t = \Delta x = .1$ implicit method

x	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.
$u_{.4}^{1}$	-.122924	-.227537	-.315474	-.388279	-.447356	-.493943	-.559081	-.553596	-.568087	-.572917
exact $u_{.4}$	-.129825	-.241132	-.335366	-.413902	-.478006	-.528808	-.567273	-.594177	-.610088	-.615353
$\lvert\varepsilon\rvert$	.006901	.013595	.019892	.25623	030650	.034865	.038192	.040581	.042001	.042435
corr. $u_{.4}$	-.130159	-.241630	-.335877	-.414342	-.478356	-.529089	-.567522	-.594432	-.610377	-.615631
$\lvert\varepsilon\rvert$	.000334	.000498	.000511	.000440	.000350	.000281	.000249	.000255	.000289	.000278

TABLE 4

$u_t + uu_x = u_{xx}$

Initial conditions $u = 1 \quad x < \frac{1}{2}$; $u = \frac{1}{2} \quad x = \frac{1}{2}$; $u = 0 \quad \frac{1}{2} < x < 1$

$\Delta x = .1 \qquad \Delta t = .01$

x	0	.1	.2	.3	.4	.5	.6	.7	.8	.9	1
corr. $u_{t=.56}$	1.	.922804	.846759	.771742	.696021	.617454	.533447	.440410	.332443	.197465	0
exact $u_{t=.56}$	1.	.926129	.848440	.772698	.696503	.617471	.532889	.439036	.329782	.192572	0
$\lvert\varepsilon\rvert$	0	.003325	.001681	.000956	.000482	.000017	.000558	.001374	.002661	.004893	0

Table 2 demonstrates that the correction procedure can easily be applied to equations with a "source" term. In the diffusion equation (67), the only coefficient or correction factor that needs to be calculated is the ratio $\frac{\alpha}{\gamma}$. For an equation with an additional (possibly constant) source term, then as well as the ratio $\frac{\alpha}{\gamma}$, the actual value of γ is needed.We have used γ for the correction to the time derivative and α for the coefficient of the space derivative. Once again, it can be seen from line 5 that the solution after applying the procedure is correct to order $\tau^4 + h^4$.

Table 3 is included in order to show that this method is also applicable to equations with derivative (Neumann) boundary conditions. The maximum of the absolute value of the error does not occur at the boundary, and the "corrected" solution is correct to fourth order in h, even at the boundaries.

As a slightly more complicated example, Table 4 shows some results for the non linear Burgers equation. Since most non-linear problems are solved by iteration, the fact that our method might require three of four iterations is no hindrance. The two types of iteration are carried out at the same time.

3.2 Elliptic equations

To be able to demonstrate that this procedure works equally well for elliptic partial differential equations, the Laplace's equation was solved on computational grids of various mesh-sizes. Some results for this are presented in Table 5. These

T A B L E 5

Solution of Laplace's equation

$u_{xx} + u_{yy} = 0$; exact solution $u = x^4 + y^4 - 6x^2y^2$

Number of grid points	Average error	Number of iterations	Time required
100 × 100	.3306	250 relaxation	120 seconds
50 × 50	.001864	250 relaxation	33 seconds
10 × 10	.001216	33 relaxation	9 seconds
	.000186	80 correction	
20 × 20	.00034	114 relaxation	20 seconds
	.000036	100 correction	
30 × 30	.00018	229 relaxation	44 seconds
	.000048	100 correction	

results suggest that the method is at its best when solving elliptic equations. This may be explained by the fact that these are solved by iteration, and, as was pointed out above, the iterations can be combined.

The explanation of Table 5 is as follows. For a grid of 100 points square on the unit square, that is, a mesh size of 0.01, after 250 iterations of the successive over relaxation process, the solution showed a very large average error. (By average error we mean $\{ \sum_{i}^{N} \sum_{j}^{N} (u_{ex} - u_{ij}) \} / N^2$). For a coarser grid of 50 points square, after 250 iterations of the SOR process, the average error was 0.001864 and 33 CPU seconds (on a CDC 6500 computer) were required. Finally, for a very coarse grid, of only 10 points square, 33 iterations of SOR yielded a reasonably accurate solution, and a further 80 iterations using the higher order method gave an average error of 0.000186, and needed only 9 CPU seconds. In other words, by using this correction procedure, the accuracy can be improved by an order of accuracy in one quarter of the time. Other results are included for grids of 20 × 20 and 30 × 30 points. Obviously, whichever method is chosen, the finer the mesh the finer the resolution of the solution. However, if the corrections are applied then the solution is more accurate yet.

3.3 Hyperbolic equations

As an example of a hyperbolic partial differential equation consider the "non viscous" form of the Burgers' equation. It is most often with hyperbolic equations that auxiliary boundary

conditions have to be invented. The differential equation is of first order, and it is rare that a first order solution will suffice. For symmetric problems, periodicity conditions can provide extra information but in general this is not the case. Since our method for higher accuracy does not require any auxiliary conditions, this problem is not encountered.

Table 6 presents the results for this procedure, the lines representing the same quantities as Table 1. It is not expected that a solution correct to 10^{-7} would be used in many circumstances, but this example serves to illustrate that the higher order method copes with all three types of partial differential equations which fact is of use when the equation can change type with time.

It is sometimes necessary to exercise a little caution when the equation or system of equations being solved is expected to produce a steady state solution, that is, when one (or more) of the derivatives may tend to zero, or if one (or more) of the derivatives passes through zero.

Equation (64) was written in such a way that we have devided throughout by γ, which is a ratio of the high and low order derivatives with respect to t. Clearly, in the steady state $\frac{\partial u}{\partial t} = 0$, and therefore γ is a ratio of zero to zero, and division by zero must be avoided. However, if we have reached a position in time where $\frac{\partial u}{\partial t} = 0$, then a steady state has been reached, and no further computation is necessary.

TABLE 6

$u_t + \sinh x \, u_x = 0$

Exact solution $u = e^{-t} \tanh \frac{x}{2}$ Initial condition $u = \tanh \frac{x}{2}$

$\Delta x = .1$ $\Delta t = .004$

x	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.
$u^1_{.036}$	.0481912	.0961415	.143616	.190388	.236246	.280995	.324458	.366486	.406943	.445777
exact $u_{.036}$	.0481919	.0961439	.143621	.190396	.236258	.281012	.324481	.366514	.406981	.445777
$\lvert\varepsilon\rvert \times 10^7$	7	24	50	80	120	70	230	280	380	0
corr. $u_{.036}$	.0481919	.0961438	.143621	.190397	.236259	.281013	.324482	.366518	.406980	.445777
$\lvert\varepsilon\rvert \times 10^7$	0	1	2	3	5	7	8	25	10	0

In the case where a spatial derivative has a zero, and calculation is required, no modification of the general procedure is needed. Consider, for example, the simple first order partial differential equation

$$\frac{\partial u}{\partial x} - \frac{\partial u}{\partial y} = x - y - \frac{5}{4} \tag{73}$$

with boundary conditions

$$\begin{aligned} u(x=0) &= \frac{1}{4}\,(2y^2 - 3y) \\ u(y=0) &= \frac{1}{2}\,x\,(x - 1) \end{aligned} \tag{74}$$

This has as exact solution

$$u = \frac{1}{2}\,x\,(x - 1) + \frac{1}{4}\,(2y^2 + 3y) \tag{75}$$

and as derivatives

$$\frac{\partial u}{\partial x} = x - \frac{1}{2}$$

$$\frac{\partial u}{\partial y} = y + \frac{3}{4}$$

Clearly, if we arbitrarily write our partial difference equation in the difference form

$$\frac{\alpha}{h}\,(u_{i+ij} - u_{ij}) - \frac{\beta}{h}\,(u_{ij} - u_{ij-1}) = x - y - \frac{5}{4}$$

or, on simplifying

$$u_{i+ij} = u_{ij} + \frac{\beta}{\alpha}(u_{ij} - u_{ij-1}) + \frac{h}{\alpha}(x - y - \frac{5}{4}) \qquad (76)$$

then solving as a first order method, with $\alpha = \beta = 1$, presents no problem. Taking $h = 0.1$, the values obtained by the first three time-like steps (in the y direction) are given in Table 7. As we know the solution, it can be seen that this formally first order solution is also exact. However, if we now were to try to obtain a better solution by the calculation of the coefficients α and β then the numerator of α will have a zero at $x = 0.5$. It was stated above that no modification of the basic method is necessary. Clearly, if during the course of calculation a position is reached where $\frac{\partial u}{\partial x} = 0$, then we will again have zero divided by zero. On inspection of β and evaluation of the right hand side of the equation (73), it can be seen that the equation is satisfied identically, and, consequently, no improvement can be expected. Therefore, calculation should proceed to the next step. This is a direct result of the fact that the method is consistent (see equation (23)).

Another reason for exercising caution when applying this method might not be anticipated. It is essential to be careful when choosing a lower order scheme with which to start the procedure. For example, to solve the hyperbolic equation

$$u_t + u_x = 0 \qquad (77)$$

with $u(t=0) = x^2$, by the widely used Lax's method, gives, as difference equation

$$u_j^{n+1} = \frac{1}{2}(u_{j+1}^n + u_{j-1}^n) - \frac{\tau}{2h}(u_{j+1}^n - u_{j-1}^n) \qquad (78)$$

TABLE 7

SOLUTION OF EQUATION (73)

j	u_j (x=0.0)	u_j (x=0.1)	u_j (x=0.2)	u_j (x=0.3)	u_j (x=0.4)	u_j (x=0.5)	u_j (x=0.6)
0	0.0	-0.0450	-0.0800	-0.1050	-0.1200	-0.1250	-0.1200
1	+0.0800	+0.0350	0.0	-0.0250	-0.0400	-0.0450	-0.0400
2	+0.1700	0.1250	0.0900	0.0650	0.0500	0.0450	0.0500
3	0.2700	0.02250	0.1900	0.1650	0.1500	0.1450	0.1500

The exact solution of the differential equation (77) is

$$u = (x - t)^2 \tag{79}$$

The exact solution of the difference equation (78) is

$$u_j^n = (jh - n\tau)^2 + n(h^2 - \tau^2) \tag{80}$$

$$= (x - t)^2 + n(h^2 - \tau^2) \tag{81}$$

It can be shown that, at a given position x, the differential equation possesses a decreasing (with t) solution, while it is quite possible for the corresponding difference equation to have a solution which increases with time. For example, with $h = 0.1$ and $\tau = 0.01$, the following three phenomena can be seen from Table 8. Firstly, at $j = 1$, the true solution is a decreasing solution and the numerical solution is increasing; secondly, at $j = 5$, the true solution decreases while the numerical solution remains constant; and thirdly, for j greater than 5 the numerical solution behaves like the actual solution. Consequently, when a value for the derivative $\frac{\partial u}{\partial t}$ is calculated for a second approximation to the solution, this can be positive, zero or negative ! This can lead to some very unexpected results.

3.4 Summary

The three sections of this chapter have shown that the higher order method is equally applicable to any of the three different types of equations, and indeed also to nonlinear or mixed type equations.

TABLE 8

SOLUTION BY LAX'S METHOD COMPARED WITH EXACT SOLUTION

j	0	1	2	3	4	5	6	7	8	9	10
u_j^0	0.0	.01	.04	.09	.16	.25	.36	.49	.64	.81	1.00
u(t=.01)	.0001	.0081	.0361	.0841	.1521	.2401	.3481	.4761	.6241	.7921	.9801
u_j^1	.0100	.0180	.0460	.0940	.1620	.2500	.3580	.4860	.6340	.8020	.9900

Since in real life the differential equations encountered are most often (highly) non linear, and possibly of a type which changes as the solution progresses, an iteration approach such as this one brings an immediate benefit. For example, solving the equation

$$u_t + uu_x = \nu u_{xx}$$

would normally be accomplished by means of the following difference equation :

$$u_j^{n+1} = u_j^n - \frac{\tau}{2h} u_j^n (u_{j+1}^n - u_{j-1}^n) + \nu \frac{\tau}{h^2} (u_{j+1}^n - 2u_j^n + u_{j-1}^n)$$

The coefficient of the first difference u is here evaluated at the old time level and a completely explicit scheme used. The use of an implicit method would imply solving a non linear system of algebraic equations which would require much iteration for even a low order solution. This explicit formula is not applied at the boundaries, the boundary conditions are inserted after the calculation of the remainder of the field. Not only would our higher order method use the existing iteration procedure to increase the accuracy of the solution but also, it would use the boundary conditions as part of the solution process.

4. FAST METHODS

4.0 Discussion

Although section 3 dealt with a method in terms of a high order of accuracy, it should be pointed out that there is always a balance between the number of points used (proportional to the amount of computational work needed) and the accuracy attained. For example, a so called second order method on the interval (0,1) with ten points will give an accuracy (normalised) of 0.01, with twenty points .0025, with 100 points .0001, etc. Therefore, instead of calling our method a higher order method, it could be called a method using less points or less work. That is to say, if no more accuracy is wanted, the same solution can be obtained with much less effort.

It was shown in section 3.3 that it is possible to solve numerically hyperbolic partial differential systems to any order of accuracy. However, the time development of hyperbolic equations is very sensitive to boundary conditions, and therefore even more sensitive to the auxiliary (artificial) conditions. It is sometimes the case that the exact time history development of the solution is not of particular interest, but only the steady state solution (if one exists). Then solution with a very dissipative numerical scheme such as Lax's is possibly an advantage, since the numerical viscosity which it injects adds stability. In this way, it reduces the dramatic effect of slightly inaccurate boundary conditions. It is relatively easy

to construct schemes which add in a numerical viscosity but lead to the correct steady state solution as $t \to \infty$.

This artificial damping will hold the solution from 'blowing up' but will not necessarily lead to any acceleration in the convergence to a steady state. With this in mind we now consider methods where accuracy is not of prime importance in the time development, but only the final, time independent, solution is of interest. By their very nature, hyperbolic equations have no inherent damping. An equation of the form of equation (77) has one characteristic direction, namely at 45 degrees to the x axis. Any singularity is propagated along this characteristic with neither attenuation nor growth. Also, in the case of equation (77) there are two relevant conditions, an initial condition and ONE boundary condition at the minimum value of x.

By solving equation (77) numerically, any singularity in the initial conditions can be attenuated, amplified or simply propagated into the field. With Lax's scheme, for example, taking equal step size in the two direction leads, in theory, to a solution which neither decays nor grows. In practise, due to computer round-off,'equal' step sizes may well lead to a diverging solution. Taking the time like step size greater than the space like step violates the CFL condition, leading to an unstable solution.

In the next section a method is presented which both complements and supplements the method of section 2. It is complementary in that it provides no time accuracy at all, and

supplementary in that the two methods used together can give a very rapid convergence to a very accurate steady state. (Using the two methods together gives a solution which is accurate in terms of a transformed time variable). Alternatively, the higher order methods can be included as the steady state is reached. Again the method is best described by means of a simple example, where, because the example is linear, detailed convergence rate and stability analyses can be carried through to find the optimum conditions to give the most rapid convergence to the steady state.

The Navier-Stokes equations of fluid dynamics can be approximated to the boundary layer equations near a solid/fluid interface, and to the Euler equations in the flow regime far from any interface. Much work has been done in the field of the boundary layer equations (see for example the method presented by Hill (19) in 1974), and so the next section concentrates on the inviscid Euler equations.

The steady two- or three-dimensional Euler equations can be solved as a hyperbolic problem in one space variable with respect to the other space variable(s). The time dependent problem can also be solved as a hyperbolic problem in time with respect to space. The method presented here aims at avoiding the spatial solution by taking it as a pseudo time dependent problem and achieving a steady state.

5. A RELAXATION METHOD FOR HYPERBOLIC EQUATIONS

5.0 Introduction

The hyperbolic system of partial differential equations which is encountered most frequently in fluid dynamics is the system describing inviscid fluid flow, known as the Euler equations. For two spatial dimensions; the Euler equations are written as follows :

$$\rho_t + (\rho u)_x + (\rho v)_y = 0$$

$$(\rho u)_t + (\rho u^2+p)_x + (\rho uv)_y = 0$$

$$(\rho v)_t + (\rho uv)_x + (\rho v^2+p)_y = 0$$

$$E_t + [u(E+p)]_x + [v(E+p)]_y = 0$$

being a continuity equation, two momentum equations and an energy equation, respectively.

The system can also be written in a shorter "vector" form, which in two dimensions looks like

$$\frac{\partial W}{\partial t} + \frac{\partial F}{\partial x} + \frac{\partial G}{\partial y} = 0$$

where

$$W = \begin{bmatrix} \rho \\ \rho u \\ \rho v \\ E \end{bmatrix} \qquad F = \begin{bmatrix} \rho u \\ \rho u^2+p \\ \rho uv \\ u(E+p) \end{bmatrix} \qquad G = \begin{bmatrix} \rho v \\ \rho uv \\ \rho v^2+p \\ v(E+p) \end{bmatrix}$$

Indeed, this vector form can also be used to describe the viscous Navier-Stokes equations if F and G contain some extra terms.

In a steady state, the system of equations can be solved as a system which is

1) hyperbolic if the flow is supersonic,
2) parabolic if the flow is sonic, and
3) elliptic if the flow is subsonic.

This fact makes the number of boundary conditions to be applied an unknown function of the solution, which is not a very desirable situation.

A further disadvantage of solving the time independent equations is met in cases where there are pockets of reversed flow. A reverse flow region will occur in corners in the boundary geometry, or other such places where there is an adverse pressure gradient. Solving a subsonic problem without reversed flow requires physically only boundary conditions from upstream whereas solving with some reversed flow will necessitate some condition(s) from downstream in addition. In other words, the problem of where to apply the boundary conditions can only be solved after the flow direction is known and of course, the flow direction is a function of the boundary conditions.

These problems can be avoided by using the time dependent system of equations.

5.1 Relaxation applied to differential equations

The basis of this acceleration procedure is best illustrated by means of an example in one space dimension and one time dimension. Therefore, consider the first approximation to the Euler equations, namely the acoustic equations :

$$\rho_t + u_x = 0$$

$$u_t + \rho_x = 0.$$

These first order equations can be rewritten as second order equations, namely

$$\rho_{tt} - \rho_{xx} = 0$$

$$u_{tt} - u_{xx} = 0.$$

Both forms of these equations show that the system is hyperbolic, since a wavelike solution exists of the form

$$\rho = \rho_0 \, e^{i\sigma(x \pm t)}$$

$$u = u_0 \, e^{i\sigma(x \pm t)} .$$

In such a solution, as t tends to infinity, the wavelike nature of the solution is preserved. However, the reason for solving time dependent equations (in this context) is to find a steady state.

By solving a differential system which has no inherent attenuation with a numerical scheme with an 'artificial viscosity' then a steady state can be found. For example, the solution of the hyperbolic equation

$$u_t + Uu_x = 0$$

by Lax's method is equivalent to solving the diffusion-like equation

$$u_t + Uu_x = \alpha_e u_{xx}$$

which has introduced an effective artificial diffusion coefficient

$$\alpha_e = \frac{\Delta x^2}{2\Delta t} - \frac{\Delta t}{2} U^2$$

Because the equation being solved is now parabolic and no longer hyperbolic in nature, it is possible to find a steady state solution as t tends to infinity. This steady state solution is the solution of the ordinary differential equation

$$Uu_x = \alpha_e u_{xx}$$

and therefore not the solution of the original hyperbolic equation.

Such dissipative schemes are widely used to find solutions to hyperbolic systems, yet obviously there is no guarantee that the steady states of the two (differential and difference) systems will coincide. Also, although convergence to some steady state is assured, the rate of convergence might well require a large amount of computational work. (The 'steady state' of such a simple equation as above is somewhat trivial). The method here described not only guarantees convergence, but at a very fast rate. The convergence rate of any iterative system can be characterised by the coefficient of t appearing in the exponential function. Three basic cases exist :

1. a purely imaginary coefficient causes any disturbance to propagate with neither damping nor growth;
2. a positive real part of the coefficient amplifies any 'error' causing the solution to diverge;
3. a negative real part of the coefficient attenuates any 'error' and causes the solution to converge, the larger the coefficient the faster the convergence.

Rather than implicitly modify the differential equation by using a dissipative numerical scheme, it is proposed to modify the differential system explicitly in order to add some sort of damping term. The rate of damping is described by the coefficient of t appearing in the exponential function, as above. By definition, for hyperbolic systems, this coefficient is imaginary, being the root of a quadratic equation. If another equation and another dependent variable are introduced, the resulting equation is cubic, and therefore has three roots. If the coefficients of the cubic are all real, then at least one of the roots will be real, and the other two either real or complex.

Since in a physical problem it is the pressure term which reduces a fluid to its mechanical equilibrium state, an artificial pressure term is introduced here in order to provide a third equation which will lead to an attenuation of wave-like disturbances. The rate at which this equilibrium state is reached can easily be determined from a Fourier type analysis of the solution.

The simple hyperbolic system

$$\rho_t + u_x = g(x)$$

$$u_t + \rho_x = 0$$

has a steady state solution :

$$u = \int g \, dx$$

$$\rho_x = 0.$$

By replacing ρ in the second equation by a pressure-like variable q, the rate at which this steady state is attained can be improved. An additional equation for this new variable q must now be introduced, such that $q \to \rho$ as $t \to \infty$.

Consider the system

$$\rho_t + u_x = g(x)$$

$$u_t + q_x = 0$$

$$q_t + \lambda u_x + \frac{1}{\tau} (q-\rho) = \lambda g(x)$$

where τ has the dimension of time.
The role of τ can most easily be seen by considering this third equation alone, with u constant with respect to x and ρ constant with respect to t. Then the third equation can be reduced to

$$q_t + \frac{1}{\tau} (q-\rho) = \lambda g(x)$$

This ordinary differential equation has the solution

$$q = A\, e^{-\frac{t}{\tau}} + \rho + \tau\lambda g(x)$$

and so, in this case, the larger the value of τ the faster the rate of convergence will be. The effect will be the same in the full system, but it will not always be possible to derive a differential equation with analytic solution.

Returning to the three equation system, a similar analysis can be executed. By cross differentiation, this system can be reduced to a single equation in q, as follows :

$$\left(q_{tt} - \lambda q_{xx}\right)_t + \frac{1}{\tau}\left(q_{tt} - q_{xx}\right) = 0.$$

Then by writing q in Fourier components as

$$q = q_0 e^{\sigma t} e^{i\alpha x}$$

the following equation is found

$$\sigma(\sigma^2 + \lambda\alpha^2) + \frac{1}{\tau}(\sigma^2 + \alpha^2) = 0$$

From the three equation system it can immediately be seen that the steady state solution of this system is identical to that of the two equation system. This formulation can be made slightly more general by including a convective term in the 'residual' $(q-\rho)$ in the third equation, by writing

$$q_t + \lambda u_x + \mu(q - \rho)_x + \frac{1}{\tau}(q - \rho) = \lambda g$$

the characteristic polynomial becomes

$$\sigma(\sigma^2 + \lambda\alpha^2) + i\alpha\mu(\sigma^2 + \alpha^2) + \frac{1}{\tau}(\sigma^2 + \alpha^2) = 0$$

By defining some new variables this equation can be simplified. Let k be a 'stretching factor', and let

$$\omega = \frac{1}{\tau k} + i\mu$$

$$z = \frac{\sigma}{k}$$

then the polynomial becomes

$$z^3 + \omega z^2 + \lambda z + \omega = 0.$$

Since it is a cubic equation, this polynomial possesses three roots. These three roots can be all real, or one real and two complex, if ω is real, or all three complex if ω is complex. For stability, it is required that the real part of z is equal to (neutral stability) or less than zero. This can be shown to be the case if $\lambda \geq 1$ and $\omega_r > 0$.

Consider first the case where there is one real root and a complex conjugate pair, i.e., $\omega_i = 0$. Let these roots be z_1 and $z_r \pm z_i$. Then, by the conditions on the coefficients of a cubic equation

$$z_1 + 2z_r = -\omega$$

$$2z_1 z_r + z_r^2 + z_i^2 = \lambda$$

$$z_1(z_r^2 + z_i^2) = -\omega$$

The cubic can be rewritten as

$$z = -\omega \frac{z^2+1}{z^2+\lambda}$$

from which it can be seen that $|z|$ is less than ω if $\lambda \geq 1$. Assume $\omega_r > 0$. By the third equation z_1 shows the same sign as $-\omega$, and hence $z_1 < 0$. Then $2z_r = -\omega - z_1$, giving

$z_r < 0$ if $|z| < \omega$, which yields

$\lambda \geq 1$.

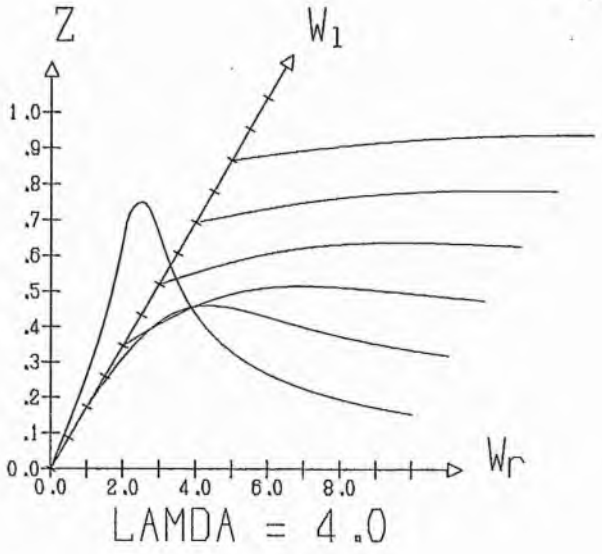
A similar analysis applies when all three roots are real, or all three are complex.

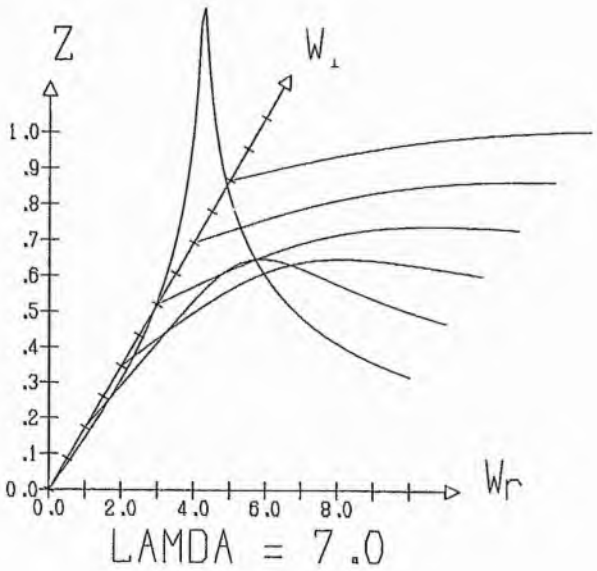
Figures 3 and 4 show the behaviour of the roots of the cubic polynomial. Since it is only the real part of the root which plays any part in the convergence (or stability) of the solution, the absolute value of the real part of z is plotted for various λ and ω.

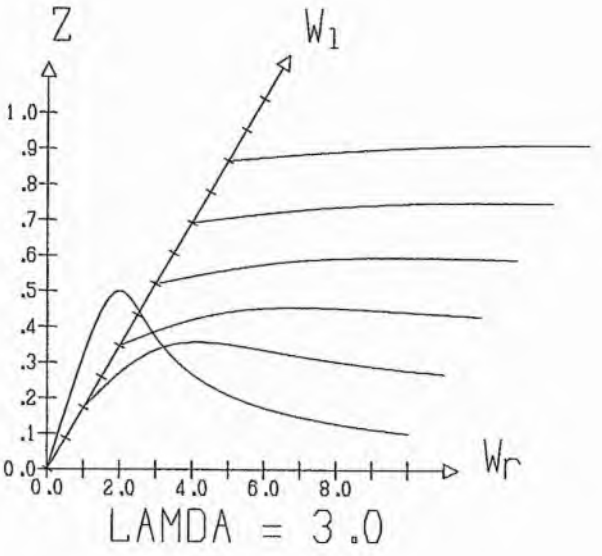
Figure 3 shows some three dimensional plots of z plotted against the real part and the imaginary part of ω for various values of λ. The missing case of $\lambda = 1$ results in two imaginary roots and a third root equal to $-\omega$.

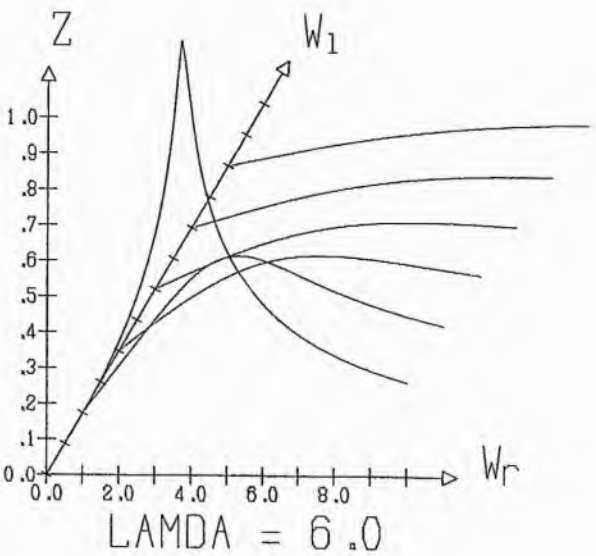
Since only the root with smallest real part determines the rate of convergence if the solution, figure 3, presents only the smallest root. For $\lambda \leq 5$, the smallest root is the real root, and so the curves are smooth. For $\lambda \geq 5$, the smallest root is initially the real root as ω increases from zero, then the real part of the complex root becomes the smaller where $\omega = 3\sqrt{\frac{\lambda-3}{2}}$.

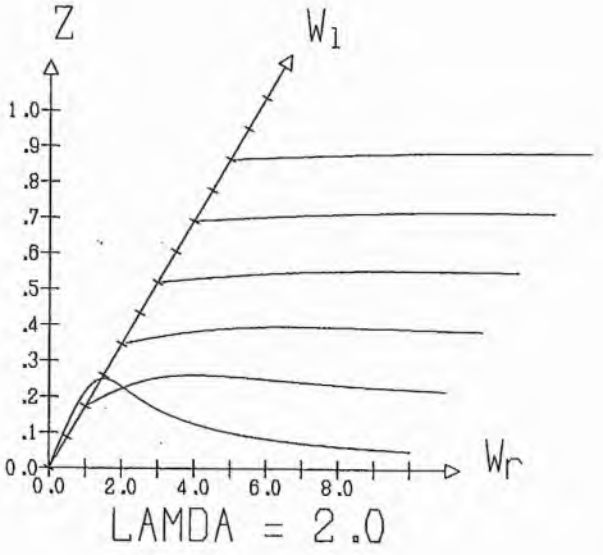
By the symmetry of the problem, it can at once be seen that the roots are symmetric with respect to the imaginary part of ω about $\omega_i = 0$. For large values of ω_i, the real parts of all three roots tend to zero and therefore there is only interest in the range $0 \leq \omega_i \leq 10$.


Z
W1
Wr
1.0
.9
.8
.7
.6
.5
.4
.3
.2
.1
0.0
0.0
2.0
4.0
6.0
8.0
LAMDA = 4.0


Z
W1
Wr
1.0
.9
.8
.7
.6
.5
.4
.3
.2
.1
0.0
0.0
2.0
4.0
6.0
8.0
LAMDA = 7.0


Z
W1
Wr
1.0
.9
.8
.7
.6
.5
.4
.3
.2
.1
0.0
0.0
2.0
4.0
6.0
8.0
LAMDA = 3.0


Z
W1
Wr
1.0
.9
.8
.7
.6
.5
.4
.3
.2
.1
0.0
0.0
2.0
4.0
6.0
8.0
LAMDA = 6.0


Z
W1
Wr
1.0
.9
.8
.7
.6
.5
.4
.3
.2
.1
0.0
0.0
2.0
4.0
6.0
8.0
LAMDA = 2.0

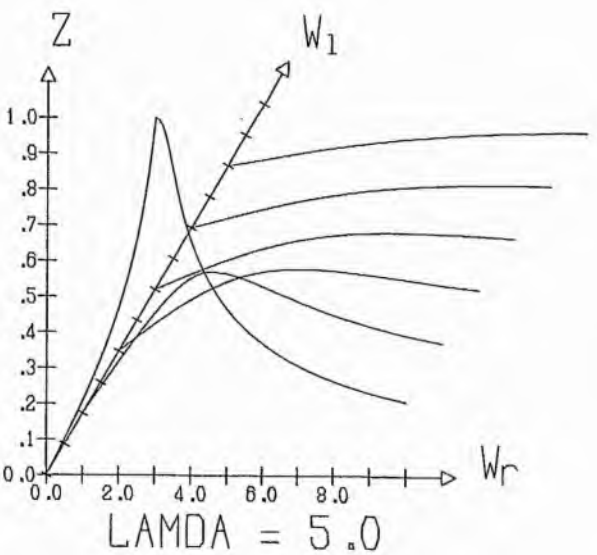
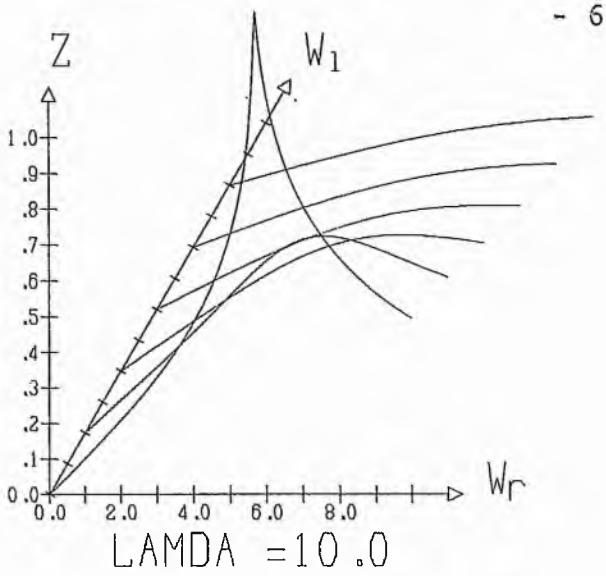
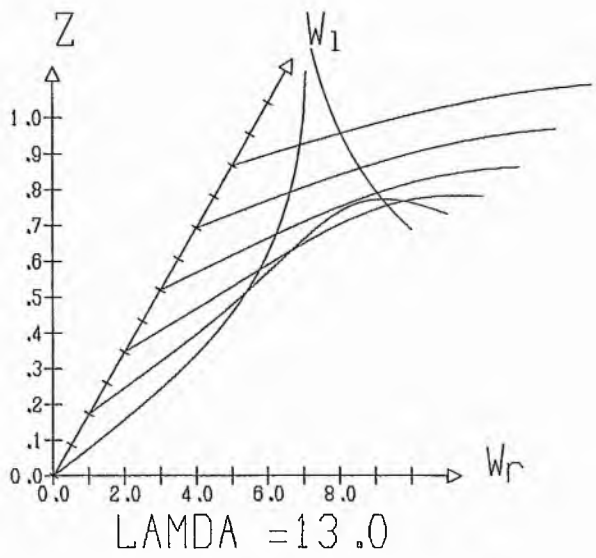
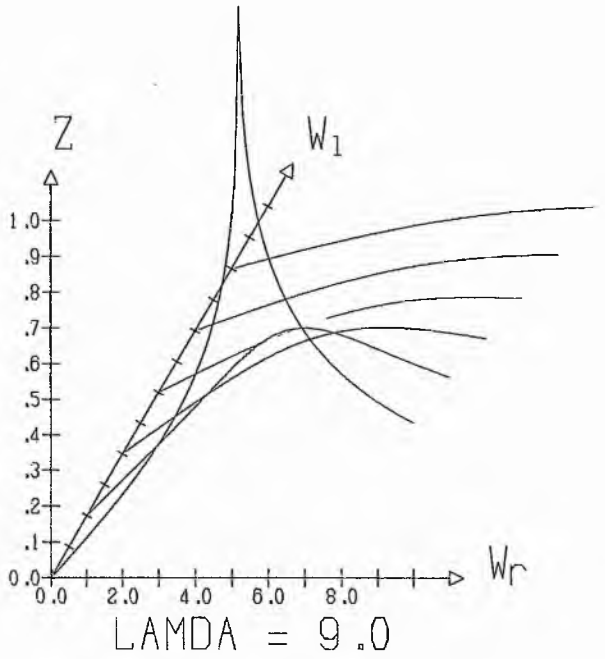
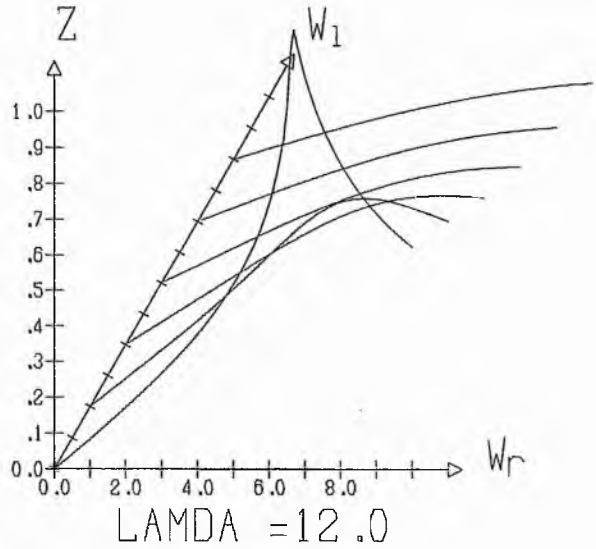

Z
W1
Wr
1.0
.9
.8
.7
.6
.5
.4
.3
.2
.1
0.0
0.0
2.0
4.0
6.0
8.0
LAMDA = 5.0

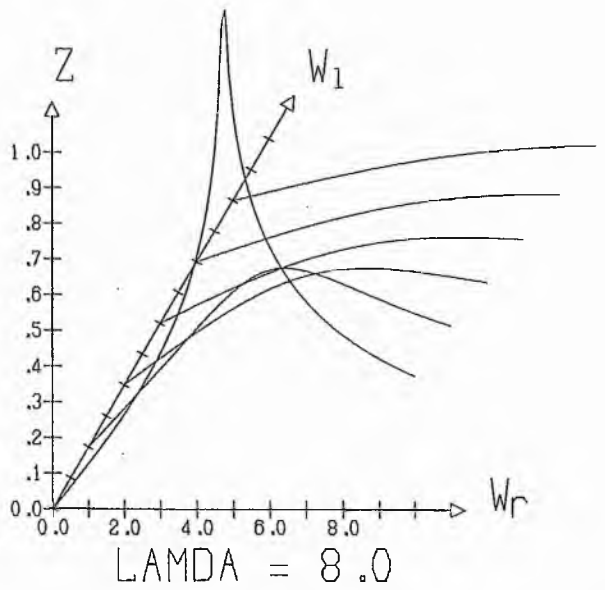
FIGURE 3a


Z
W1
Wr
1.0
.9
.8
.7
.6
.5
.4
.3
.2
.1
0.0
0.0
2.0
4.0
6.0
8.0
LAMDA =10.0


Z
W1
Wr
1.0
.9
.8
.7
.6
.5
.4
.3
.2
.1
0.0
0.0
2.0
4.0
6.0
8.0
LAMDA =13.0


Z
W1
Wr
1.0
.9
.8
.7
.6
.5
.4
.3
.2
.1
0.0
0.0
2.0
4.0
6.0
8.0
LAMDA = 9.0


Z
W1
Wr
1.0
.9
.8
.7
.6
.5
.4
.3
.2
.1
0.0
0.0
2.0
4.0
6.0
8.0
LAMDA =12.0


Z
W1
Wr
1.0
.9
.8
.7
.6
.5
.4
.3
.2
.1
0.0
0.0
2.0
4.0
6.0
8.0
LAMDA = 8.0

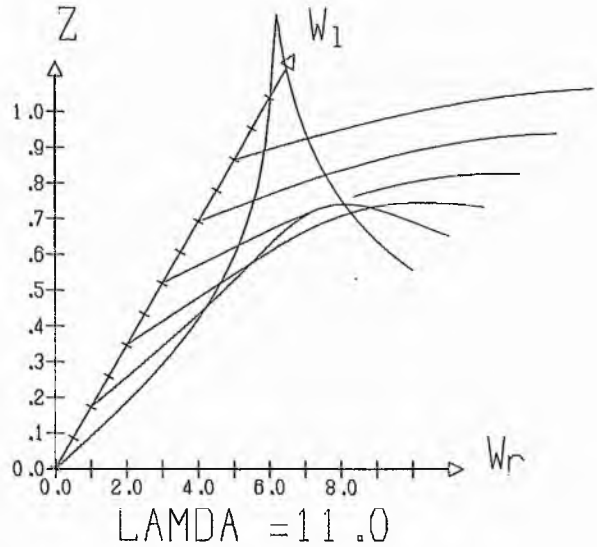
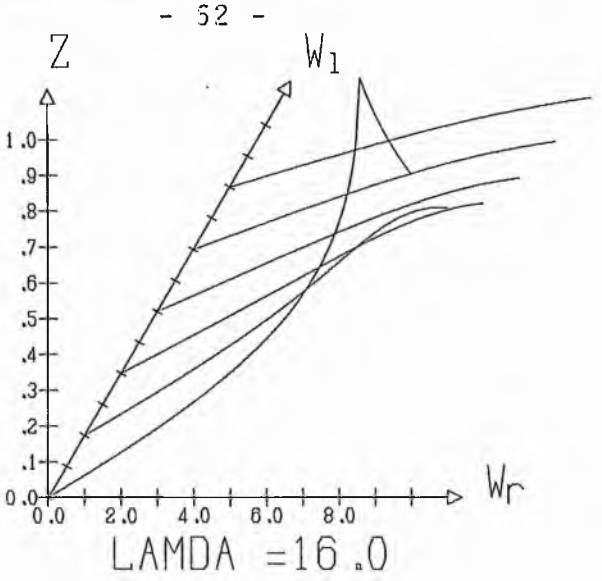
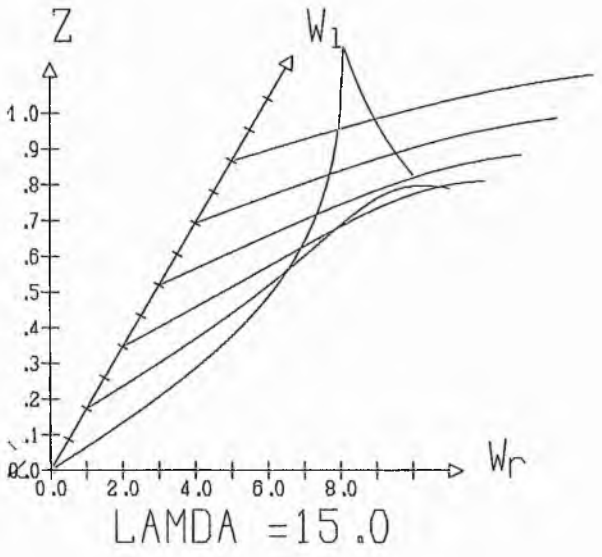

Z
W1
Wr
1.0
.9
.8
.7
.6
.5
.4
.3
.2
.1
0.0
0.0
2.0
4.0
6.0
8.0
LAMDA =11.0

FIGURE 3b





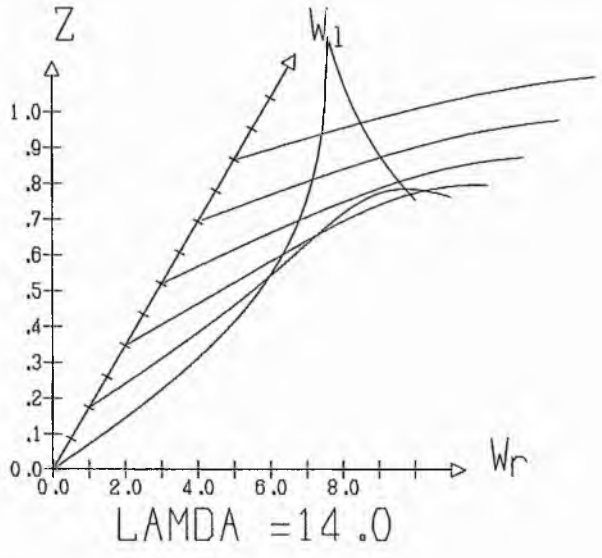


FIGURE 3c

11.02.09
13/08/76
750 SEC
181 REC

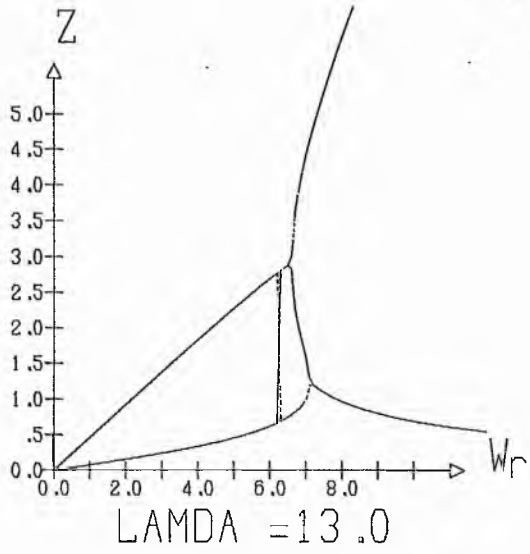
Figure 3 shows clearly that the maximum real part with respect to the imaginary part of ω is always at the place where this disappears, that is, where ω is real. The figure also shows that the maximum of z with respect to the real part of ω is at $\lambda = 9$, where $\omega = 3\sqrt{3}$ and $z = \sqrt{3}$.

Figure 4 plots the real part of all three roots for various values of λ. Since from figure 3 it was noticed that the maximum of z is where ω is real, only the real parts have been plotted.

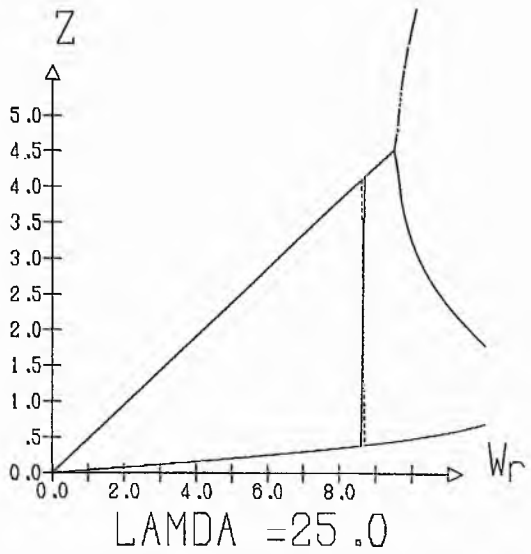
The plot for $\lambda = 5$ shows that the roots intercept at the maximum of the one root.

The plot for $\lambda = 9$ does not show very clearly the intersection due to the fact that the on-line plotter has a finite accuracy, and the intersection occurs at irrational values, namely $\omega = 3\sqrt{3}$ and $z = \sqrt{3}$.

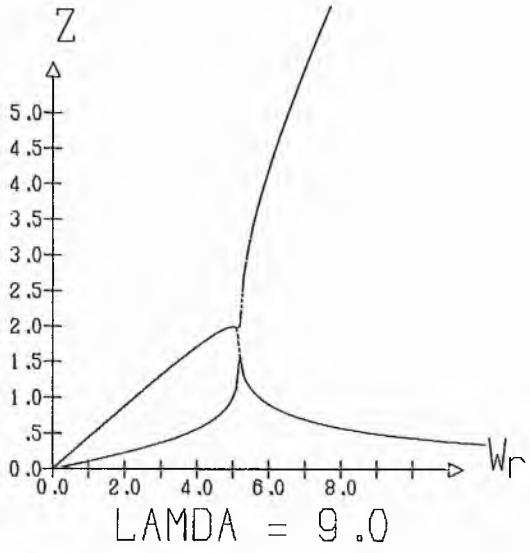
The spurious vertical lines appearing in the plots for $\lambda \geq 13$ are due to the fact that the roots are found by an iterative (Newton's) method, and are found in a different order for different values of ω. For these cases where $\lambda < 9$, it can immediately be seen that there are only two real parts, in other words the roots are one real and a complex conjugate pair, whereas for $\lambda > 9$ there are certain values of ω which give rise to the situation where there can be three distinct roots, all of which are real. For the most rapid convergence the largest value of the smallest root z is needed. This occurs where the roots coincide, which is at $z = \sqrt{3}$, $\omega = 3\sqrt{3}$, $\lambda = 9$.



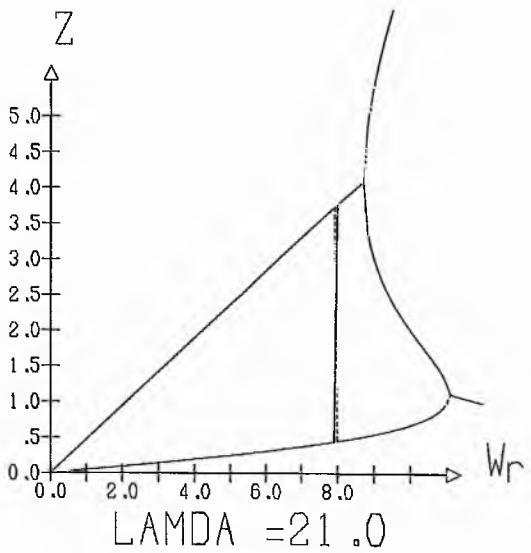
LAMDA =13.0



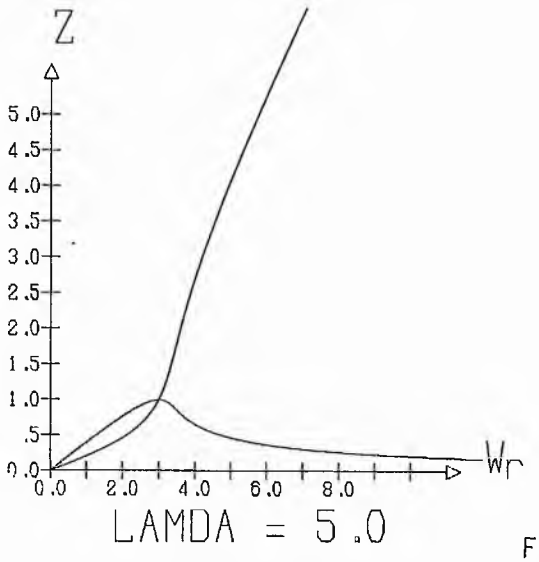
LAMDA =25.0



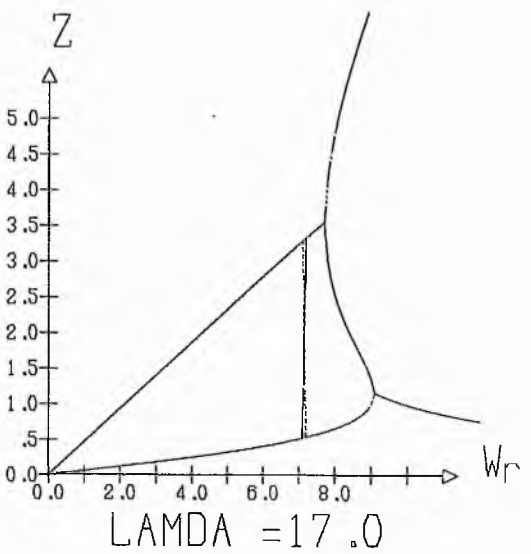
LAMDA = 9.0



LAMDA =21.0



LAMDA = 5.0



LAMDA =17.0

FIGURE 4

Since by adding one equation to the system, the characteristic polynomial is now capable of possessing all real roots in place of all imaginary roots before,in effect the system has been transformed from a hyperbolic system into an elliptic one, in the transient (time dependent) solution of the problem. Naturally, it is now possible to apply any difference scheme to the solution of this system in order to find the required steady state solution - which is now certain to exist. The higher order method described in sections 2 and 3 above is ideal, in that the differential formulation contains the necessary damping terms. As with other schemes, it will be shown in what follows that the rate of convergence can further be improved (or indeed worsened) by suitable choice of difference scheme. Many difference schemes will lead to exactly the same characteristic polynomial developed here for the differential case. This means that the fact that the optimum conditions have been found for the roots of this equation,can be used to optimise the various difference schemes.

A more complicated system of equations can even be shown to lead to this same cubic characteristic polynomial. The three dimensional 'acoustic' equations can be reduced to this equation as follows.

Consider the five equation system

$$\rho_t + u_x + v_y + w_z = 0$$

$$u_t + q_x = 0$$

$$v_t + q_y = 0$$

$$w_t + q_z = 0$$

$$q_t + \lambda(u_x + v_y + w_z) = -\frac{1}{\tau}(q - \rho)$$

The corresponding characteristic polynomial is then the result of substituting the relevant Fourier components into the equation

$$\left(q_{tt} - \nabla^2 q\right)_t + \frac{1}{\tau}\left(q_{tt} - \nabla^2 q\right) = 0$$

which with

$$q = q_0\, e^{\sigma t}\, e^{i\alpha x}\, e^{i\beta y}\, e^{i\gamma z}$$

gives

$$\sigma\left[\sigma^2 + \lambda(\alpha^2+\beta^2+\gamma^2)\right] + \frac{1}{\tau}\left[\sigma^2 + (\gamma^2+\beta^2+\gamma^2)\right] = 0$$

Defining

$$\omega = \frac{1}{\tau(\alpha^2+\beta^2+\gamma^2)^{1/2}}$$

$$z = \frac{\sigma}{\sqrt{(\alpha^2+\beta^2+\gamma^2)}}$$

gives again

$$z(z^2 + \lambda) + \omega(z^2 + 1) = 0$$

This is exactly the same equation as that derived above for the one space dimension system. Therefore, the same results presented in the graphs of figures 3 and 4 can be applied, and we can restrict ourselves to the one dimensional form of the equations.

Since the new variable q has been introduced as a pressure-like variable, some information must also be provided concerning this variable. The only assumption made so far is that the steady state condition is such that $q \equiv \rho$; that is, that the residual $(q - \rho)$ tends to zero. The most rational condition to

apply with regard to initial and boundary conditions is to commence with no residual and also to impose this criterion on the boundary.

Starting with no residual, that is with $q \equiv \rho$ and taking $\lambda = 1$, the solution obtained is identical to the solution of only the original hyperbolic system. This is then a useful reference case for difference schemes. With dissipative numerical schemes, where a steady state solution can be found (through a method which damps any initial disturbance), setting $\lambda = 1$ will yield the number of iterations which would normally be required. Not only can this method be used with computational schemes which add no artificial viscosity term (either explicitly or implicitly), but also it can add stability to procedures which would otherwise be unstable. Starting the computations from initial conditions with the residual $(q - \rho)$ different from zero will immediately yield a non zero residual and may sometimes accelerate the convergence in the initial stages.

It can be seen that the stability and convergence rate optimisation analysis based on the differential problem is also based on the largest wave number present in the Fourier component. In other words, to damp each wave number, there exists a different set of optimum parameters. Fortunately, the various values needed in order to attenuate the various wave numbers are usually very close to each other. In fact, most physical problems of fluid flow will have a dominant wave number and it is therefore this wave number which would be used in approximating an optimum set of parameters. It will be seen in the next section that the iteration or transient problem is not particularly sensitive to

a small deviation away from the optimum in the free parameters. This is a very useful fact when it comes to solving nonlinear problems, because in such problems the stability and also the rate of convergence are functions of the unknown solution.

5.2 Relaxation applied to difference schemes

When we now attempt to apply some difference schemes to the solution of our quasi-elliptic scheme, some more parameters are introduced. In addition to the parameters λ and ω, the ratio of the time step to each space step size also appears. The problem of of stability and the related problem of convergence rate are then functions of (at least) three independent variables. As for differential problems, it is often possible in the difference system to optimise the values taken by the free parameters λ and ω, and therefore find the values which will require the least number of iterations. Indeed, it is frequently the case that the optimum values for the difference case are close to those found in the differential formulation, namely $\lambda \simeq 9$ and $\omega \simeq 5$.

Although the system of equations being solved now contains one extra equation, the small amount of extra (programming) work involved is easily justified. It will be shown that, even with a rough approximation to the best values of the coefficients λ and ω, a gain in the number of iterations of at least an order of magnitude is achieved. For a system of five equations such as the Euler system, adding one further equation is not as significant in that it only adds one fifth to the total amount of work, yet adds in both convergence rate and in stability.

By means of both of these two methods used together, a highly accurate steady state solution can be achieved very rapidly. Use of a first order difference method with the higher order method provides a stable scheme which will need no auxiliary boundary conditions.

Before discussing some difference schemes in greater detail, some variables are introduced in order to simplify the notation. Let

$t_n = n\,\Delta t$ for $n = 0,1, \ldots$

$x_j = jh$ for $j = 0,1, \ldots,N$

$r = \frac{\Delta t}{h}$

$s = \frac{\Delta t}{\tau}$ $\qquad \rho_j^n = \rho\,(t_n,x_j)$

$u_j^n = u(t_n,x_j)$ $\qquad q_j^n = q\,(t_n,x_j)$

5.2.1 First order differencing

Since the Euler system to be solved is a first order system, then the most logical place to start is with a first order difference scheme. The system of equations

$$\rho_t + u_x = g(x)$$

$$u_t + q_x = 0$$

$$q_t + \lambda u_x = \lambda g(x) - \frac{1}{\tau}(q - \rho)$$

then becomes a difference system as follows :

$$\rho_j^{n+1} = \rho_j^n - r(u_j^n - u_{j-1}^n) + \Delta t\, g$$

$$u_j^{n+1} = u_j^n - r(q_j^n - q_{j-1}^n)$$

$$q_j^{n+1} = q_j^n - \lambda r(u_j^n - u_{j-1}^n) + \lambda\Delta t\, g - s(q_j^n - \rho_j^n)$$

If now the dependent variables are written in terms of Fourier components, a stability analysis, and simultaneously a convergence-rate analysis, can be performed. Let

$$\rho = \rho_0\, \xi^n\, e^{i\beta hj} \qquad u = u_0\, \xi^n\, e^{i\beta hj}$$

$$q = q_0\, \xi^n\, e^{i\beta hj}$$

then the following algebraic system of equations can be solved. In this case it is not the solution of this system which is of interest, but only the case of when the solution exists. To put this another way, the coefficients ρ_0, u_0, q_0 are not required from the stability point of view, but merely the conditions under which the solution can be solved.

$$\rho_0\, \xi^{n+1} e^{i\beta hj} = \rho_0\, \xi^n e^{i\beta hj} - r(u_0\, \xi^n e^{i\beta hj} - u_0\, \xi^n e^{i\beta h(j-1)}) + \Delta t\, g$$

$$u_0\, \xi^{n+1} e^{i\beta hj} = u_0\, \xi^n e^{i\beta hj} - r(q_0\, \xi^n e^{i\beta hj} - q_0\, \xi^n e^{i\beta h(j-1)})$$

$$q_0\, \xi^{n+1} e^{i\beta hj} = q_0\, \xi^n e^{i\beta hj} - \lambda r(u_0 \xi^n e^{i\beta hj} - u_0\, \xi^n e^{i\beta h(j-1)}) + \lambda\Delta t\, g - s(q_0\, \xi^n e^{i\beta hj} - \rho_0\, \xi^n e^{i\beta hj})$$

In the homogeneous case ($g = 0$) the factor $\xi^n e^{i\beta hj}$ can be divided throughout giving

$$\rho_0(\xi - 1) + ru_0 (1 - e^{-i\beta h}) = 0$$

$$u_0(\xi - 1) + rq_0 (1 - e^{-i\beta h}) = 0$$

$$q_0(\xi - 1 + s) + r\lambda u_0(1 - e^{-i\beta h}) - s\rho_0 = 0$$

Let

$$\eta = \xi - 1$$

$$i\theta = 1 - e^{-i\beta h}$$

Then

$$\rho_0\eta + ir\theta u_0 = 0$$

$$u_0\eta + ir\theta q_0 = 0$$

$$q_0(\eta + s) + ir\theta\lambda u_0 - s\rho_0 = 0$$

In matrix form, this is

$$\begin{bmatrix} \eta & ir & 0 \\ 0 & \eta & ir\theta \\ -s & ir\theta\lambda & \eta+s \end{bmatrix} \begin{bmatrix} \rho_0 \\ u_0 \\ q_0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Then the condition that this holds for all ρ_0, u_0, q_0 is

$$\text{Det} \begin{bmatrix} \eta & ir\theta & 0 \\ 0 & \eta & ir\theta \\ -s & ir\theta\lambda & \eta+s \end{bmatrix} = 0,$$

or

$$\begin{vmatrix} \eta & ir\theta & 0 \\ 0 & \eta & ir\theta \\ -s & ir\theta\lambda & \eta+s \end{vmatrix} = 0$$

This determinant can be expanded, giving

$$\eta\left[\eta(\eta+s) + \lambda r^2\theta^2\right] + sr^2\theta^2 = 0$$

Let $z = \frac{\eta}{r\theta}$, $\omega = \frac{s}{r\theta}$

then this becomes

$$z^3 + \omega z^2 + \lambda z + \omega = 0$$

This is exactly the same equation as was derived above for the continuous (differential) form. The optimum values, and the values for stability are immediately available. For stability we need $|\xi| \leq 1$, that is to say $|\eta + 1| \leq 1$.

This implies $0 \leq \eta \leq 2$

Then taking these two limits separately, the following results can be derived :

a) For $0 \leq \eta$, we find $0 \leq zr\theta$

which is the case if $\lambda \geq 1$.

b) For $\eta \leq 2$, we find $zr\theta \leq 2$

which can be considered as a stability bound on r, that is

$$r \leq \frac{2}{z\theta}.$$

The optimum conditions, $\lambda = 9$, $\omega = 3\sqrt{3}$ such that $z = \sqrt{3}$ imply

$$s = 3\sqrt{3}\, r\theta.$$

As for the mone common successive over relaxation methods used for elliptic equations, the rate of convergence can be defined uniquely by means of the so called "Reciprocal rate of convergence". This is defined as

$$R = \frac{-1}{\log|\xi|} .$$

Taking the logarithm to base 10 gives a measure of the number of iterations required to gain each correct decimal digit. Clearly

$$R_{10} \equiv \frac{-1}{\log_{10}|\xi|} = A\, R_e$$

where $R_e \equiv \frac{-1}{\ell n|\xi|}$

and $A = \log_{10} e.$

In the non relaxed two equation system (retrieved by setting $\lambda=1$ and $q\equiv\rho$), the roots of the cubic in z are $z = 0$ and $\pm i$. This gives

$\xi = zr\theta + 1 = 1 \pm ir\theta$ and $r\theta$.

and only the complex conjugate pair of roots is of interest, since starting with $q = \rho$ eliminates the third root. Expanding the exponential in θ gives

$$i\theta = 1 - e^{-i\beta h}$$
$$= 1 - (1+i\beta h+O(h^2))$$
$$= i\beta h + O(h^2)$$

Thus $\xi = 1 \pm ir\beta h + O(h^2)$

$$|\xi|^2 = 1 + (\pm r\beta h)^2$$
$$= 1 + r^2\beta^2h^2$$

$$Re = \frac{-2}{\ell n(1-r^2\beta^2h^2)} = \frac{2}{r^2\beta^2h^2}$$

Thus, for an accuracy of 10^{-4} on the interval [0,1], 10^4 and points would be needed, and 10^8 iterations per decimal digit. This is clearly out of the question in most cases, since at one millisecond per iteration this would require several hours of computer time.

5.2.2 Lax's method

A more commonly used method for hyperbolic systems is the one named after Lax, and described above in Section 3.3 This method is found more frequently in the literature because it injects an artificial (or numerical) viscosity which causes the solution to converge more rapidly to a steady state solution, without the restriction of requiring too much extra computational effort. It is formally correct to first order in the time-step-size, and correct to second order in the spatial step-size.

Consider again the 'acoustic' equations, given by

$$\rho_t + u_x = g(x)$$

$$u_t + q_x = 0$$

$$q_t + \lambda u_x = \lambda g(x) - \frac{1}{\tau}(q-\rho).$$

The equivalent difference scheme for Lax's method is as follows :

$$\rho_j^{n+1} = \frac{1}{2}(\rho_{j+1}^{n} + \rho_{j-1}^{n}) - \frac{1}{2} r(u_{j+1}^{n} - u_{j-1}^{n}) + \Delta t\, g$$

$$u_j^{n+1} = \frac{1}{2}(u_{j+1}^{n} + u_{j-1}^{n}) - \frac{1}{2} r(q_{j+1}^{n} - q_{j-1}^{n})$$

$$q_j^{n+1} = \frac{1}{2}(q_{j+1}^n + q_{j-1}^n) - \frac{1}{2} r\lambda(u_{j+1}^n - u_{j-1}^n) + \lambda\Delta t\, g - s(q_j^n - \rho_j^n)$$

Following the same procedure as above in order to determine the stability limits and the corresponding rate at which the solution converges, the relevant matrix equation is

$$\begin{bmatrix} \xi-\cos\beta h & ir\sin\beta h & 0 \\ 0 & \xi-\cos\beta h & ir\sin\beta h \\ -s & ir\lambda\sin\beta h & \xi-\cos\beta h+s \end{bmatrix} \begin{bmatrix} \rho_0 \\ u_0 \\ q_0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

On equating the determinant of the matrix to zero and simplifying it is found that

$$\eta^3 + s\eta^2 + \lambda r^2\sin^2\beta h\eta + sr^2\sin^2\beta h = 0$$

where $\eta = \xi - \cos\beta h$.

Setting $z = \frac{\eta}{r\sin\beta h}$, $\omega = \frac{s}{r\sin\beta h}$ gives

$$z^3 + \omega z^2 + \lambda z + \omega = 0$$

Once again this is exactly the same equation as was derived above for the differential version of the equations. Therefore, the figures 3 and 4 can be used to determined the optimum conditions. As a base for comparison, with no relaxation (that is with $\lambda = 1$) the two roots are

$$z = \pm i$$

giving

$$\xi = \pm i\, r\sin\beta h + \cos\beta h$$

so $|\xi|^2 = \cos^2\beta h + r^2\sin^2\beta h$

$= 1 + r^2\beta^2h^2 + O(\beta^3h^3)$

and $R_e = \frac{2}{r^2\beta^2} h^{-2}$

For $1 < \lambda < 9$, z has one real value and a complex conjugate pair. Let $z = \alpha, \delta \pm i\varepsilon$. Then for $z = \alpha$:

$$\xi = \alpha r \sin\beta h + \cos\beta h$$

$$= 1 + \alpha r \beta h + O(h^2)$$

and for $z = \delta \pm i\varepsilon$

$$\xi = (\delta \pm i\varepsilon)\, r \sin\beta h + \cos\beta h$$

$$= r\sin\beta h + \cos\beta h \pm i\varepsilon r \sin\beta h$$

$$|\xi|^2 = \left(\delta r \sin\beta h + \cos\beta h\right)^2 + \varepsilon^2 r^2 \sin^2\beta h$$

Thus $R_e \propto h^{-1}$.

Therefore it can be seen that this form of relaxation gives, for Lax's scheme, a reciprocal rate of convergence proportional to h^{-1} instead of the normal h^{-2}. For h smaller than 0.1 this is certainly a significant advantage. Not only at the optimum values, but also for any $\lambda > 1$ and $\omega > 0$ (or $s > 0$) the necessary number of iterations is greatly reduced.

For the differential version of the equations, it can easily be seen that the optimum values to take for the free parameters are given by a straightforward condition. Since there are three roots to the cubic characteristic equation, and these three roots appear in an exponential function, the best values will be obtained where these three roots are equal. To be equal, the three roots must all be real. Then, since the sum of the roots is given by minus the coefficient of z-squared, namely, $-\omega$,

$3z = -\omega$

Also, from the conditions on the coefficients of the cubic equation,

$3z^2 = \lambda$

$z^3 = -\omega$

and hence, $z^2 = 3$

and $z = \pm\sqrt{3}$ at $\omega = \pm 3\sqrt{3}$ and $\lambda = 9$.

These values are shown in Figs. 3 and 4.

On the other hand, the values of interest for the difference version of the equations are slightly different. The aim now is to minimise the values taken by ξ, the amplification factor. The variable to be taken into consideration is then the modulus of ξ since it is this which determines the convergence. The conditions for when the three values coincide can also be calculated as follows. Let the three roots in terms of z be

$z = \alpha$ and $z = \delta \pm i\varepsilon$

Then, in terms of ξ this gives

$\xi = \alpha r \sin\beta h + \cos\beta h$ (real)

$\xi = (\delta \pm i\varepsilon)\, r \sin\beta h + \cos\beta h$ (complex)

In terms of $|\xi|^2$, this latter expression becomes

$$|\xi|^2 = [\delta r \sin\beta h + \cos\beta h]^2 + \varepsilon^2 r^2 \sin^2\beta h$$

Using the conditions for the roots of a cubic, then

$$\alpha + 2\delta = -\omega$$

$$2\alpha\delta + \delta^2 + \varepsilon^2 = \lambda$$

$$\alpha(\delta^2+\varepsilon^2) = -\omega\sigma$$

Equating the two values for $|\xi|^2$ gives the optimum values for ω and λ. Thus

$$[\alpha r \sin\beta h + \cos\beta h]^2 = [\delta r \sin\beta h + \cos\beta h]^2 + \varepsilon^2 r^2 \sin^2\beta h$$

$$\alpha^2 r^2 \sin^2\beta h + 2\alpha r \sin\beta h \cos\beta h + \cos^2\beta h$$
$$= \delta^2 r^2 \sin^2\beta h + 2\delta r \sin\beta h \cos\beta h + \cos^2\beta h + \varepsilon^2 r^2 \sin^2\beta h$$

If we ignore the "small" terms in $\sin^2\beta h$ then this will reduce to

$$2\alpha r \sin\beta h \cos\beta h = 2\delta r \sin\beta h \cos\beta h$$

or, to first order in βh

$$\alpha = \delta = -\frac{\omega}{3}$$

which will also give $\alpha = \pm\sqrt{3}$ $\omega = \pm 3\sqrt{3}$ $\lambda = 9.$

A higher order value can be achieved if necessary but a trial and error procedure will probably be found to have the same results when more complicated problems are being solved and often the difference equations are not as easy to analyse.
These results are substantiated when the method is used on a computer.

5.2.3 The Lax-Wendroff method

As has been demonstrated above, the Lax's method gives to a scheme an artificial viscosity which gives guaranteed convergence, but at the expense of accuracy. A more accurate procedure is the Lax-Wendroff scheme, which is based on the Taylor series expansion in time. For example, the simple two equation system

$$\rho_t + u_x = 0$$

$$u_t + \rho_x = 0$$

can be written as

$$\rho^{n+1} = \rho^{n} + \Delta t\, \rho_t + \frac{1}{2} \Delta t^2\, \rho_{tt} + O(\Delta t^3)$$

$$u^{n+1} = u^{n} + \Delta t\, u_t + \frac{1}{2} \Delta t^2\, u_{tt} + O(\Delta t^3)$$

The first order derivatives with respect to time can be replaced by spatial derivatives (with respect to x) immediately on using the definition of the system, and the second derivatives by simple cross differentiation giving

$$\rho_{tt} + u_{xt} = 0$$

$$u_{xt} + \rho_{xx} = 0$$

$$\Rightarrow \rho_{tt} = \rho_{xx}, \qquad u_{tt} = u_{xx}.$$

The Taylor series for ρ then becomes

$$\rho^{n+1} = \rho^{n} - \Delta t\, u_x + \frac{1}{2} \Delta t^2\, \rho_{xx} + O(\Delta t^3)$$

Naturally, this process can be continued if desired, since $\rho_{ttt} = -u_{xxx}$ and all the further time derivatives can be evaluated in terms of spatial derivatives. The accuracy of the solution therefore rests solely on the accuracy of the spatial derivatives. There are many ways in which this method can be programmed as a numerical scheme for a computer. For example, the spatial derivatives can be evaluated at either the old or the new time level. MacCormack uses a two step scheme which avoids the analytic differentiation of the system, which is useful when the functions appearing are non linear.

The reciprocal rate of convergence has been introduced above as a measure of the value of a particular method, but this is a variable normally associated with elliptic systems. Therefore, before considering the relaxation form of the equations, the reciprocal rate of convergence is derived for the commonly used Lax-Wendroff scheme, applied to the two equation system above.

Using second order differencing in both time and space gives

$$\rho_j^{n+1} = \rho_j^n - \frac{1}{2} r \delta_x u_j^n + \frac{1}{2} r^2 \delta_x^2 \rho_j^n$$

$$u_j^{n+1} = u_j^n - \frac{1}{2} r \delta_x \rho_j^n + \frac{1}{2} r^2 \delta_x^2 u_j^n$$

In terms of Fourier components, if

$$\begin{pmatrix} \rho^n \\ u^n \end{pmatrix} = \begin{pmatrix} \rho_0 \\ u_0 \end{pmatrix} \xi^n e^{i\beta hj}$$

then

$$\delta_x e^{i\beta hj} = 2i \sin\beta h \, e^{i\beta hj}$$

$$\delta_x^2 e^{i\beta hj} = 2(\cos\beta h - 1) \, e^{i\beta hj}$$

giving

$$\begin{bmatrix} \xi - 1 + r^2(1-\cos\beta h) & ir \sin\beta h \\ ir \sin\beta h & \xi - 1 + r^2(1-\cos\beta h) \end{bmatrix} \begin{bmatrix} \rho_0 \\ u_0 \end{bmatrix} = 0$$

Setting the determinant to zero gives

$$\left[\xi - 1 + r^2(1-\cos\beta h)\right]^2 + r^2 \sin^2\beta h = 0$$

$$\xi = 1 - r^2(1-\cos\beta h) \pm ir \sin\beta h$$

$$|\xi|^2 = \left(1 - r^2(1-\cos\beta h)\right)^2 + r^2 \sin^2\beta h$$

$$= 1 - 2r^2(1-\cos\beta h) + r^4(1-\cos\beta h)^2 + r^2 \sin^2\beta h$$

Expanding

$$\cos\beta h = 1 - \frac{1}{2}\beta^2 h^2 + \frac{\beta^4 h^4}{24} + O(h^6)$$

$$\sin\beta h = \beta h - \frac{\beta^3 h^3}{6} + O(h^5)$$

$$\sin^2\beta h = \beta^2 h^2 + \frac{\beta^4 h^4}{3} + O(h^6)$$

then, correct to $O(h^6)$

$$|\xi|^2 = 1 - 2r^2\left(\frac{\beta^2 h^2}{2} - \frac{\beta^4 h^4}{24}\right) + r^4\left(\frac{\beta^2 h^2}{2} - \frac{\beta^4 h^4}{24}\right)^2 + r^2\left(\beta^2 h^2 - \frac{\beta^4 h^4}{3}\right)$$

$$= 1 - r^2\left(\beta^2 h^2 - \frac{\beta^4 h^4}{12} - \beta^2 h^2 + \frac{\beta^4 h^4}{3}\right) + r^4 \frac{\beta^4 h^4}{4}$$

$$|\xi|^2 = 1 + r^2\beta^4 n^4 \left[\frac{1}{12} - \frac{1}{3}\right] + r^4 \frac{\beta^4 h^4}{4}$$

$$= 1 - \frac{1}{4} r^2(1-r^2)\ \beta^4 h^4$$

and hence

$$R = \frac{8}{r^2(1-r^2)}\ \beta^{-4}\ h^{-4}$$

Thus the normal form of the Lax-Wendroff method is two orders of magnitude slower in convergence than the Lax method.

For a given h, the rate of convergence can easily be optimised setting $\frac{\partial R}{\partial r} = 0$. Since

$$\frac{\partial R}{\partial r} = 8\beta^{-4} h^{-4} (r^2-r^4)^{-2} (2r-4r^3) = 0$$

this gives three values for r, namely $r = 0, \pm \frac{1}{\sqrt{2}}$.
It is commonly assumed in the literature that the best convergence rate is obtained close to the stability bound ($r = 1$), but this is clearly not the case, ($r = 0$ is no convergence, and $r = \frac{1}{\sqrt{2}}$ is optimum).

For the three equation system

$$\rho_t + u_x = 0$$

$$u_t + q_x = 0$$

$$q_t + \lambda u_x + \frac{1}{\tau} (q-\rho) = 0$$

the analysis is very similar. The Taylor series are

$$\rho^{n+1} = \rho^n - \Delta t\ u_x + \frac{1}{2} \Delta t^2 q_{xx}$$

$$u^{n+1} = u^{n} - \Delta t\, q_x + \frac{1}{2}\Delta t^2\left[\lambda u_{xx} + \frac{1}{\tau}(q-\rho)_x\right]$$

$$q^{n+1} = q^{n} - \Delta t\left[\lambda u_x + \frac{1}{\tau}(q-\rho)\right] + \frac{1}{2}\Delta t^2\left[\lambda q_{xx} + \frac{1}{\tau}(\lambda-1)u_x + \frac{1}{\tau^2}(q-\rho)\right]$$

which leads then to the system

$$\begin{bmatrix} \xi-1 & ir\sin\beta h & r^2(1-\cos\beta h) \\ \frac{1}{2}ir s\sin\beta h & \xi-1-\lambda r^2(\cos\beta h-1) & ir\sin\beta h-\frac{1}{2}ire\sin\beta h \\ \frac{1}{2}s^2-s & i\lambda r\sin\beta h-\frac{1}{2}s(\lambda-1)ir\sin\beta h & \xi-1+s-\lambda r^2(\cos\beta h-1)-\frac{1}{2}s^2 \end{bmatrix}\begin{bmatrix} \rho_0 \\ u_0 \\ q_0 \end{bmatrix} = 0$$

The analysis can be continued as before and the analysis follows the same lines. However, a better method is presented in the next section.

5.2.4 "Euler" differencing

A method of differencing which is usually disregarded as being unstable, but which is formally second order accurate in the space-step-size is known as Euler differencing. The difference scheme obtained by this means is the following

$$\rho_j^{n+1} = \rho_j^{n} - \frac{1}{2} r\,(u_{j+1}^{n} - u_{j-1}^{n})$$

$$u_j^{n+1} = u_j^{n} - \frac{1}{2} r(q_{j+1}^{n} - q_{j-1}^{n})$$

$$q_j^{n+1} = q_j^{n} - \frac{1}{2}\lambda r\,(u_{j+1}^{n} - u_{j-1}^{n}) - s(q_j^{n} - \rho_j^{n})$$

Taking the Fourier components, by writing $\rho = \rho_0 \xi^n e^{i\beta hj}$, and similarly for u and q, gives

$$\begin{vmatrix} \xi - 1 & ir\sin\beta h & 0 \\ 0 & \xi - 1 & ir\sin\beta h \\ -s & i\lambda r\sin\beta h & \xi - 1 + s \end{vmatrix} = 0$$

as a determinant equation to solve. Setting

$\eta = \xi - 1, \qquad m = r\sin\beta h$

gives

$$\begin{vmatrix} \eta & im & 0 \\ 0 & \eta & im \\ -s & i\lambda m & \eta+s \end{vmatrix} = 0$$

and

$\eta^3 + s\eta^2 + \lambda m^2\eta + m^2 s = 0$

Now, let

$\eta = r\sin\beta h \quad z = mz$

$s = \omega m$

then $\qquad z^3 + \omega z^2 + \lambda z + \omega = 0.$

Once again, this is the same equation as for the differential systems, and the same analysis applies.
At the optimum for the differential system,

$z = -\sqrt{3} \qquad \omega = -3\sqrt{3} \qquad \lambda = 9$

so $\quad \eta = -r\sin\beta h \cdot \sqrt{3}$

and $\quad \xi = 1 + \eta = 1 - r\sin\beta h \cdot \sqrt{3}$

Clearly $|\xi| < 1$ so long as

$r\sqrt{3}\ \sin\beta h < 2$

Expanding $\sin\beta h = \beta h + O(\beta^6 h^6)$

$\xi = 1 - r\sqrt{3}\ \beta h + O(\beta^6 h^6)$

and the rate of convergence is better than that for the Lax-Wendroff method, and the coding required is simpler.

5.2.5 A semi-implicit scheme

An obvious enhancement to the scheme described in section 5.2.4 is to write the third equation in an implicit manner. That is, let us write our system as

$$\rho_j^{n+1} = \rho_j^n - \frac{1}{2} r\,(u_{j+1}^n - u_{j-1}^n)$$

$$u_j^{n+1} = u_j^n - \frac{1}{2} r\,(q_{j+1}^n - q_{j-1}^n)$$

$$q_j^{n+1} = q_j^n - \frac{1}{2} \lambda r (u_{j+1}^{n+1} - u_{j-1}^{n+1}) - s(q_j^{n+1} - \rho_j^{n+1}).$$

Following the same procedure as before gives

$$\begin{vmatrix} \xi - 1 & ir\sin\beta h & 0 \\ 0 & \xi - 1 & ir\sin\beta h \\ -s\xi & i\xi r\lambda \sin\beta h & \xi - 1 + s\xi \end{vmatrix} = 0$$

Let $\eta = \xi - 1$ and $m = r\sin\beta h$, then

$$\begin{vmatrix} \eta & im & 0 \\ 0 & \eta & im \\ -s(1+\eta) & (1+\eta)i\lambda m & \eta+s(1+\eta) \end{vmatrix} = 0$$

giving

$$(1+s)\eta^3 + (s+m^2\lambda)\eta^2 + (\lambda m^2+sm^2)\eta + sm^2 = 0$$

or, with $s = \omega m$, $\eta = mz$:

$$(1+s)z^3 + (\omega + \frac{m}{\omega}\lambda)z^2 + (\lambda+s)z + \omega = 0, \qquad \text{or}$$

$$(1+\omega m)z^3 + (\omega + \frac{m}{\omega}\lambda)z^2 + (\lambda+\omega m)z + \omega = 0$$

Clearly, as $s \to 0$ this becomes our familiar equation

$$z^3 + \omega z^2 + \lambda z + \Omega = 0$$

However, the equation possesses too many unknowns (m, ω, λ) to be analysed usefully. The place where the three roots are equal can be determined as follows.

For the equation :

$$x^3 - ax^2 + bx - c = 0$$

this happens where

$$D = Q^2 + P^3 = 0$$

where

$$Q = -a^3 + \frac{9}{2}ab - \frac{27}{2}c$$

$$P = -a^2 + 3b$$

Taken individually

$$P = 0 \quad \rightarrow \quad a^2 = 3b$$

$$Q = 0 \quad \rightarrow \quad a^3 = \frac{9}{2} ab - \frac{27}{2} c$$

$$\rightarrow \quad ab = 9c$$

Returning to our equation,

$$(1+\omega m)z^3 + (\omega + \lambda \frac{m}{\omega})z^2 + (\lambda+\omega m) + \omega = 0$$

$$a = - \frac{\omega + \lambda \frac{m}{\omega}}{1 + \omega m} \qquad b = \frac{\lambda + \omega m}{1 + \omega m} \qquad c = \frac{-\omega}{1 + \omega m}$$

Then $a^2 = 3b \rightarrow \left[\frac{\omega + \frac{\lambda m}{\omega}}{1 + \omega m}\right]^2 = 3 \frac{\lambda + \omega m}{1 + \omega m}$

and $ab = 9c \rightarrow (\omega + \lambda \frac{m}{})(\lambda + \omega m) = 9\omega(1 + \omega m).$

These two equations can then be solved simultaneously for various values of s as shown in the table below.

Computer experiments have shown this method to work very well at any values of λ and ω in the vicinity of the optimum values, and that the predicted behaviour of the iteration procedure is indeed what happens.

s	ω	$\frac{1}{\lambda}$
10^{-6}	$\sqrt{3}$	$\frac{1}{9}$
10^{-5}	1.7320	$\frac{1}{9}$
10^{-4}	1.7318	.11110
70^{-3}	1.7297	.11100
2×10^{-3}	1.7274	.11090
4×10^{-3}	1.7228	.11068
6×10^{-3}	1.71816	.11047
8×10^{-3}	1.71363	.11026
10^{-2}	1.7090	.11005
2×10^{-2}	1.6865	.10901
4×10^{-2}	1.6430	.10697
6×10^{-2}	1.6017	.104499
8×10^{-2}	1.5621	.10308
.1	1.5243	.10122
.2	1.3579	.09272
.5	1.0144	.07318
.6	0.9336	.06813
1	0.7037	.05263

6. CONCLUSIONS

In the paper presented above two methods have been presented for use when solving by computer differential systems. On the one hand, a method was presented which can be used to great advantage when a cheap method for high accuracy is sought. The method can result in as accurate a solution as is required without the usual drawbacks of fictitious boundary conditions being needed and a much larger computer time being required. It can be used in only a part of the field of solutions if rapid changes are expected in a localised area. Equally, it can be used to give a rough idea of the solution of a differential system with much less computational effort.

On the other hand, a method has been demonstrated which dramatically increases the convergence speeds for the solution of hyperbolic systems. By adding one more equation to a system (usually of three or five equations) a much higher rate of convergence is achieved at the small expense of a little more programming. The method also adds stability to the system, a much desired side effect of any numerical method.

Both methods are fairly simple to program into existing computer programs, and with interactive methods parameters can be changed while the solution progresses to speed the convergence further.

The combination of the two methods presents a way of optimising the utilisation of computer time.

REFERENCES

1. RICHARDSON, L.F. (1910): The Approximate Arithmetical Solution by Finite Differences of Physical Problems involving Differential Equations, with an Application to the Stresses in a Masonry Dam.
Transact. Roy. Soc. London, A 210, pp 307-357.

2. COURANT, FRIEDRICHS & LEWY (1928): On the Partial Difference Equations of Mathematical Physics.
IBM Journal, March 1967, pp 215-234.

3. THOM (1933): The Flow past Circular Cylinders at Low Speeds.
Proc. Roy. Soc. London, A141, pp 651-666.

4. SHORTLEY & WELLER (1938): The Numerical Solution of Laplace's Equation.
J. Applied Physics, Vol. 9, pp 334-348.

5. SOUTHWELL (1946): Relaxation Methods in Theoretical Physics.
Oxford University Press.

6. FRANKEL (1950): Convergence Rates of Iterative Treatments of Partial Differential Equations.
Math. Tables and other Aids to Computation, Vol. 4, pp 65 - 75.

7. CRANK & NICOLSON (1947): A Practical Method for Numerical Evaluation of Solutions of Partial Difference Equations of the Heat Conduction Type.
Proc. Cambridge Phil. Soc., Vol. 43, No 50.

8. von NEUMANN (1944): Proposal and analysis of a numerical method for the treatment of hydrodynamical shock problems.
Nat. Def. & Res. Com. Report AM 551.

9. PEACEMAN & RACHFORD (1955): The Numerical Solution of Parabolic and Elliptic Differential Equations.
J. Soc. Indust. Applied Mathematics, Vol. 3, No 1, March.

10. DOUGLAS & RACHFORD (1956): On the Numerical Solution of Heat Conduction Problems in Two and Three Space Variables.
Transact. Am. Math. Soc., Vol. 82, pp 421-439.

11. LAX (1954): Weak Solutions of Nonlinear Hyperbolic Equations and their Numerical Computation.
Comm. Pure & Appl. Math., Vol. 7, pp 159-193.

12. LAX & WENDROFF (1960): Systems of Conservation Laws.
Comm. Pure & Appl. Math., Vol. 13, pp 217-237.

13. RICHTMYER (1963): A Survey of Difference Methods for non Steady Fluid Dynamics.
NCAR Technical Note 63-2, Boulder Colorado.

14. MacCORMACK (1969): The Effect of Viscosity in Hypervelocity Impact Cratering.
AIAA Paper No 69-354.

15. TAYLOR (1974): Numerical Methods for Predicting Subsonic, Transonic and Supersonic Flow.
AGARDograph 187.

16. HARLOW (1964): The Particle-In-Cell Computing Method for Fluid Dynamics.
Methods in Computational Physics, Vol. 3, p. 319.

17. MADER (1964): The Two Dimensional Hydrodynamic Hot Spot
LASL Report LA 3077, Los Alamos Scientific Laboratory, Los Alamos, New Mexico.

18. HILL & WIRZ (1976): Compact Higher Order Finite Difference Methods.
von Karman Institute, Internal Note 50.

19. HILL (1974): A Numerical Method for the Solution of the Two Dimensional Steady Laminar Boundary Layer Equations.
von Karman Institute, Technical Note 103.